REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
3.Nov.99

3. REPORT TYPE AND DATES COVERED
DISSERTATION

4. TITLE AND SUBTITLE
THE IMPACT OF ACCENT, NOISE, AND LINGUISTIC PREDICTABILITY ON THE INTELLIGIBILITY OF NON-NATIVE SPEAKERS OF ENGLISH

5. FUNDING NUMBERS

6. AUTHOR(S)
MAJ SCOTT KIMBERLY R

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
UNIVERSITY OF FLORIDA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
THE DEPARTMENT OF THE AIR FORCE
AFIT/CIA, BLDG 125
2950 P STREET
WPAFB OH 45433

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
FY99-401

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Unlimited distribution
In Accordance With AFI 35-205/AFIT Sup 1

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

19991117 075

14. SUBJECT TERMS

15. NUMBER OF PAGES
131

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT

18. SECURITY CLASSIFICATION OF THIS PAGE

19. SECURITY CLASSIFICATION OF ABSTRACT

20. LIMITATION OF ABSTRACT

THE IMPACT OF ACCENT, NOISE, AND LINGUISTIC PREDICTABILITY ON THE INTELLIGIBILITY OF NON-NATIVE SPEAKERS OF ENGLISH

BY

KIMBERLY R. SCOTT

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1999

ACKNOWLEDGMENTS

Special acknowledgment is given to Dr. Alice Dyson, who served as committee chair for this dissertation. Her time, guidance, patience, and encouragement are sincerely appreciated. I would also like to express my gratitude to the members of my committee-- Dr. Patricia Kricos, Dr. Howard Rothman, Dr. Christine Sapienza, and Dr. Jean Casagrande--for their guidance and encouragement.

My appreciation is extended to my colleagues in the Department of Communication Sciences and Disorders. I would especially like to thank Viviane Marino, who graciously found my non-native speakers and who patiently listened to and challenged me in endless discussions about my research. I would also like to thank Wayne King for his assistance in mixing and recording the audio tapes. My thanks to Sarah Ahmed, Ana Mendes-Schwartz, Bernadette Cesar-Lee, Jodi Brinkley, and Cynthia Core for their helpful tips, constant encouragement, and positive attitudes.

My sincere appreciation is extended to my military mentors who guided my career in such a way that I could take advantage of this opportunity. I am grateful for the financial support provided by the Air Force Institute of Technology and the United States Air Force.

I would also like to thank my friends far and near who have offered support and encouragement at every step of my studies. And last but not least, my sincere gratitude and appreciation go to the 55 volunteers who participated in this study.

TABLE OF CONTENTS

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

THE IMPACT OF ACCENT, NOISE, AND LINGUISTIC PREDICTABILITY ON THE INTELLIGIBILITY OF NON-NATIVE SPEAKERS OF ENGLISH

By

Kimberly R. Scott

August 1999

Chair: Alice M. Dyson
Major Department: Communication Sciences and Disorders

In many situations today non-native speakers of English must speak English as an international language or as a common language between two non-native speakers. Such communication is often complicated by adverse listening conditions such as noise and high stress levels. This study examined the effects of linguistic predictability and noise factors on the intelligibility of non-native speakers of English with varying degrees of accent when their listeners were native English speakers.

Speech recordings were elicited from four adult male native speakers of Brazilian Portuguese and one native speaker of English. Sentences from the *Speech Perception in Noise* lists were read by each speaker, representing native, mild, mild-moderate, moderate-strong, and strong foreign accents. Sentences were mixed with multi-talker

babble with a signal-to-noise ratio of 6 dB, 10 dB, and 15 dB. Target words in half of the sentences were highly predictable, and the remaining half were of low predictability.

All 50 listeners were native speakers of English. They wrote down the last word of each *SPIN* sentence from recordings of random selections of speakers and noise levels and rated spontaneous speech samples for degree of perceived accent and intelligibility pre- and post- *SPIN* listening task.

Analyses of the data suggest that all three factors--accent, noise, and predictability--had a combined effect on the intelligibility of non-native speakers of English. Even the intelligibility of the native speaker was compromised when the signal-to-noise ratio was low and when the linguistic predictability was also low. When the native listener was challenged further by the addition of a foreign accent, intelligibility was even more compromised. This effect was greater as the degree of accent became progressively stronger.

CHAPTER 1
INTRODUCTION AND REVIEW OF THE LITERATURE

Introduction

As the number of non-native speakers of English continues to increase, international attention has been drawn to the importance of speech intelligibility in individuals with foreign accent. It has been estimated that the number of people in the world who use English for some purpose ranges between 750 million and a billion and a half. Approximately 300 million people are native speakers of English (Strevens, 1988). The incentives for those who speak English as their second language (L2) are varied. For example, many non-native speakers are immigrants who need to speak English in order to survive in an English-speaking culture. There are also individuals whose countries recognize more than one national language. Others may be required to speak English in job-related functions in which English is recognized as the international language. Such situations include English as the international language of the business world, the air, and the sea.

Communication among native speakers and non-native speakers is often complicated when such communication takes place in less than ideal circumstances. For example, in international airspace, air traffic controllers and air crews often experience communication breakdowns as a result of accented speech, differences in phrasing, and

less than ideal radio systems used to transmit messages. A report from the National Aeronautics and Space Administration's (NASA) Aviation Safety Reporting System (ASRS, 1997) indicated that 1,804 foreign language related incidents were reported between January 1988 and December 1996. On November 16, 1996, the Associated Press reported that a collision of an airliner and a cargo plane over India was due to "heavily accented speech" resulting in a communication breakdown between pilots and air traffic controllers ("Language Barriers can Cost Lives," 1996). In such a setting, poor speech intelligibility associated with non-native speakers of English would be considered a safety hazard with human lives at risk.

The purpose of this study is to investigate particular contextual (linguistic) and environmental (noise) factors and their relationship to speech intelligibility of non-native speakers of English with varying degrees of accent. A second goal is to identify the degree to which these factors influence native listeners' perceived intelligibility of non-native speakers of English. The following review will focus upon factors influencing speech intelligibility, measures of speech intelligibility, problems with the assessment of intelligibility, the effect of accent on intelligibility, and factors influencing degree of perceived accent.

Review of the Literature

Speech Intelligibility

Speech intelligibility is a term often confused with (and sometimes used interchangeably with) speech perception. Although it may be inappropriate to use the terms interchangeably, the definitions of the two terms overlap and are often used in the

literature to mean the same thing. One reason for the confusion in terminology is that measurements of intelligibility and measurements of perception both rely on judgments made by listeners. Speech perception can be defined as "the identification of phonemes, the vowels and consonants of language, largely from acoustic cues and the recognition of phonemes in combination as a word" (Nicolosi, Harryman, & Kresheck, 1989, p. 247). The term intelligibility, on the other hand, is defined as the understandability of speech (Yorkston, Dowden, & Beukelman, as cited in Kent, 1992). Intelligibility is dependent on the intent of the speaker and the accuracy of the listener in perceiving the intended message (Munro & Derwing, 1995a; Schiavetti, 1992). Speech perception is the broader term, encompassing detection, discrimination, identification, and comprehension, in addition to intelligibility. If a message is to be intelligible, it must be perceived. However, for a sound to be perceived does not necessarily require that it be intelligible. Perception typically operates on smaller units, whereas intelligibility operates on the whole word or message.

According to Smith and Nelson (1985), judgments of intelligibility are additionally clouded by misunderstandings at the levels of comprehensibility and interpretability. They defined intelligibility as word or utterance identification and emphasized that it must be viewed separately from comprehensibility and interpretability. Comprehensibility is the recognition of the meaning of words and utterances. Comprehensibility is compromised when the listener can repeat the word or utterance but is unable to understand its meaning in the context in which it appears. Interpretability is defined as the meaning underlying the words and utterances.

Interpretability is compromised when the listener recognizes the utterance but is unable to understand the speaker's intentions behind it. It would appear that intelligibility is a vital building block; neither comprehensibility nor interpretability is possible if a word or utterance cannot be recognized.

Factors Influencing Speech Intelligibility

Intelligibility is a judgment made by a listener; *"that sentence means X"*. Factors that influence a listener's judgment of intelligibility rise from two sources defined as linguistic and non-linguistic. Linguistic factors involve matters of content, such as the level of difficulty of the message; matters of style, such as speed or hesitations; and matters of linguistic form, such as how close the form of the message is to the listener's expectations. Nonlinguistic factors include the listener's relationship with the speaker and with what the speaker is saying; physical characteristics of the speaker; distracting factors in the environment; the psychological state of the listener; the attitudes toward the native language of the speaker and listener (Fayer & Krasinski, 1987); and the acoustic aspects of the speech signal (Denes & Pinson, 1993).

Linguistic factors

Although speech perception is highly dependent upon the acoustic features of the speech wave, it is also significantly influenced by our expectations, our knowledge of the speaker, the rules of grammar, and the topic being discussed (Denes & Pinson, 1993). Context affects our perception by influencing our expectations. Sentences provide information on grammar and subject matter that lead us to what we expect to hear. Familiarity with the subject matter and articulatory peculiarities of the speaker

tend to make speech intelligible even when the acoustic cues alone, because of noise, poor articulation, or unfamiliar dialect, may be insufficient for accurate perception. One of the most important linguistic factors is that we must "know" the language to which we are listening. That is, we know its typical set of vowels and consonants, its words, and its sentence structure. For normal adults, native language perception is rapid, effortless, robust, and largely unavailable to the conscious. However, for adults listening to a non-native language, speech perception can be slow, laborious, and fragile, and it often involves conscious, analytical skills (Strange, 1997). For the native listener, then, "general context is often so compelling, we often know positively what is going to be said even before we hear the words. That is why under normal conditions, we understand speech with ease and certainty, despite the ambiguities of the acoustic cues" (Denes & Pinson, 1993, p. 183).

Cross-linguistic research over the past 25 years has clearly demonstrated that experience with a particular phonological system affects the perceptual performance of adults, especially when they are responding to native versus non-native phonetic distinctions. However, this same research has provided evidence that the effect is not the same for all listeners, nor for all phonetic distinctions, nor under all task conditions. For example, Beddor and Gottfried (1995) studied Japanese speakers' perception of the English [r] - [l] contrast and found that such non-native distinctions were easier to perceive in some syllable positions than in others. Recent emphasis has been placed on the need for understanding the precise nature of the effects of linguistic experience and how these effects come about.

Linguistic experience. Beddor and Gottfried (1995) suggested that linguistic "experience" involves a set of variables that include phonetic and phonological factors as well as the linguistic background of the listener. Research involving language-specific experience has emphasized the comparison of perception of selected phonetic distinctions across groups of listeners whose native languages differ in the use or distribution of that distinction. Voicing distinctions cued by voice onset time (VOT) are a classic example (e.g., Lisker & Abramson, 1970, as cited in Beddor & Gottfried, 1995). Other phonetic distinctions would include comparisons in which the target distinction is native for one language group and non-native for another (e.g., investigation of the [r] - [l] distinction for English and Japanese speakers mentioned above). The usual perceptual outcome of such investigations has been that listeners are good discriminators of phones that are phonemically distinct in their native language but poor discriminators of nonphonemic differences. However, it has become apparent that the differences in phonemic inventories alone are not sufficient to explain the variability in perceptual differences among listeners (Beddor & Gottfried, 1995; Strange 1995). According to Beddor and Gottfried (1995) "considerable evidence indicates that some phonetic differences are more discriminable than others, independent of their phonemic status in a language" (p. 209). For example, nonphonemic phonological factors may influence the ease or difficulty in the perception of an unfamiliar phonemic distinction. That is, if the nonphonemic feature is used to contrast other pairs of segments in the native language, listeners may perform better when that particular feature is used in the distinction of an unfamiliar segment. In English, a lengthening of the preceding vowel helps listeners

distinguish between voiced and voiceless consonants ([bæ:d, bæt]). An unfamiliar set of voiced/voiceless cognates may be distinguished by this same cue. However, studies of this point are just beginning. Polka (1992) hypothesized that Farsi speakers' experience with the velar-uvular place distinctions of the voiced stops [g] and [G] would facilitate the perception of this place distinction in other manners of articulation, but she was unable to support this hypothesis. It appears that both phonetic and phonological experience may influence the ease or difficulty of perceiving phonetic distinctions. Experience with a common phonetic feature that may be used to contrast other pairs of segments and phonetic exposure to the relevant sounds through allophonic variation in the listener's native language may facilitate the perception of new phonetic distinctions.

It has also been found that the dissimilarities between two phonemes may enhance perception (Beddor & Gottfried, 1995). For example, according to Flege's (1995) speech learning model, sounds in a second language that are dissimilar to sounds in the native language will be relatively easy to discriminate, whereas sounds that are the most difficult to learn are those that occur in the close acoustic neighborhood of the native phonemes (Kent, 1997). Kuhl's (as cited in Kuhl & Iverson, 1995) native language magnet model also states that phonetic units from a foreign language are more difficult to perceive if they are similar to a category in the adult's own native language. However, novel sounds that are not similar to a native language category are relatively easy to discriminate.

Linguistic background. The unique linguistic background of each listener is a part of his/her linguistic experience. Linguistic background variables may include the number

of other languages to which a listener has been exposed, the age of exposure to these languages, the nature of this exposure (e.g., conversational versus reading and writing experience), and the overall degree of proficiency in these languages. Research has not determined whether broader, more general linguistic experience, such as early bilingualism, influences non-native perceptual judgments. Werker (1986) found that multilingual subjects in her study did not perform better than monolingual subjects on non-native phonetic distinctions that were novel to both groups. However, in contrast, Polka (1992) found that subjects who had learned two languages as children demonstrated more native-like perception of a non-native distinction than did monolingual subjects, suggesting that language-general experience may influence perception (as cited in Beddor & Gottfried, 1995). Buus, Florentine, Scharf, and Canevet (1986) found that the degree of familiarity and the extent of exposure to English as a second language also had an effect on the ease or difficulty of understanding that language in the presence of noise. They compared 14 non-native speakers of English who were native speakers of French with 4 native speakers of American-English. The French speakers were divided according to their exposure to an American-English environment. The listening task consisted of 53 simple sentences spoken in standard American-English and presented in a background of white noise. The noise level at which the listener could repeat 50% of the sentences established the Noise Tolerance Level. Buus and colleagues concluded that as the proficiency in English increased, the ability to repeat 50% of the sentences with a higher level of background noise also increased. The difference between the Noise Tolerance Level of the listeners with

minimal exposure to English and the native American-English listeners was approximately 12 dB. As the listeners exposure to English increased, the difference between native listeners and highly proficient non-native speakers became minimal (3 dB).

Non-linguistic parameters involved in speech perception

Psychological factors. The psychological state of the listener can be a factor influencing the intelligibility of a message. Psychological variables include individual beliefs, affective states, aptitude, learning style, motivation, and personality. For example, differences in the speaker/listener culture and linguistic background can upset what might otherwise be a relatively straightforward exchange of information. Gass and Varonis (1984) examined how native speakers responded to questions of information from non-native speakers. Students enrolled in an intensive English Language Program asked strangers for directions to a train station. It was clear that native speakers responded differently to non-native speakers than they did to native speakers. In addition to echoing a part of the question asked by the non-native speaker, the native speakers exhibited a reluctance to get involved in a conversation with non-native speakers. Even a native speaker in the guise of a non-native speaker was rebuffed after asking the way to the train station. Communication between individuals of different backgrounds, whether native or non-native can be strained. Catford (1950) used the example of his familiarity with the Arabic culture which facilitated his understanding of a message that could otherwise be misunderstood. When an Arabic speaker was heard to say "Excuse me, I am now going to bray" (p. 14), Catford's experience with Arabic

speakers' [p]-[b] confusion combined with his awareness of the cultural context facilitated the intelligibility of the message. In such cases, experience and familiarity with the non-native speakers cultural background assists the native listener in the intelligibility of the message (Catford, 1950; Gass & Varonis, 1984).

Physical characteristics of the speaker also influence the listener's willingness to perceive a message that is somewhat deviant from what is expected. Judgments of intelligibility are strongly influenced by the listener's preconceived ideas about a particular non-native speaker. The personality and accent of individual non-native speakers and even the country from which they come may influence the judgment of the listener (Morley, 1993; Varonis & Gass, 1982). The reluctance of native speakers to converse with non-native speakers as seen in the Gass and Varonis (1984) study, is a clear example of the biasing effects of speaker characteristics. Several researchers (e.g., Eisenstein, 1983; Ludwig, 1982; Ryan, 1983 as cited in Anderson-Hsieh, Johnson, and Koehler, 1992) have investigated native speaker reactions to non-native speakers independent of pronunciation. In addition to influencing judgments of non-native speakers, physical appearance and history have been shown to influence even trained clinicians in their perceptual judgments of individuals with speech disorders (Kent, 1996).

Environmental Factors. Factors in the environment often influence a listener's judgment of intelligibility. Such factors include noise, limits of a transmission system, distortions, and interruptions. Early experimental studies summarized by Denes and Pinson (1993) involved the investigation of noise interference using white noise. It was

found that the impact of noise varies over a range from having no effect on speech intelligibility at a 20 dB signal-to-noise ratio to reducing word articulation scores to 50% at a 0 dB signal-to-noise ratio. A 20 dB signal-to-noise ratio represents a speech signal that is 20 dB more intense than the noise signal, and a 0 dB signal-to-noise ratio is representative of average intensities of speech and noise that are about equal. As a general rule, researchers have suggested normal conversation can occur without much difficulty at a level where a 50% word articulation score is achieved (Denes & Pinson, 1993). The word articulation score typically is determined by the percentage of words correctly identified in a list of phonetically balanced words. Although speech is often intelligible in everyday life even when its intensity is lower than that of noise, Nicolosi et al. (1989) state that a signal-to-noise ratio greater than 6 dB is needed for satisfactory communication. This may be the result of the listener's use of multiple sensory modalities, such as visual cues and nonverbal cues, in addition to the auditory signal.

Additional research involving the effect of noise on speech intelligibility has made use of multi-talker babble in an attempt to represent everyday listening situations. For example, Kalikow, Stevens, and Elliott (1977) used multi-talker babble in their design of the *Speech Perception in Noise (SPIN)* test to examine the speech perception of individuals who are hearing impaired. The babble of voices produced by several speakers has been shown to interfere with speech intelligibility more than the use of a random nonspeech noise such as the clatter of dishes, traffic and other transportation noises, office machines in operation, and ringing telephones. The babble of a few voices can produce interference that exceeds interference due solely to masking of individual

sounds. Babble noise contains false speech cues and increases the load on the attention and memory processes that are involved in understanding sentences (Kalikow et al., 1977).

Cross-linguistic researchers have studied the effect of noise on non-native listeners compared to native listeners of American English. For example, Buus et al. (1986) measured the ability of native French speakers to repeat simple American-English sentences presented in white noise. They found a 12 dB difference between the Noise Tolerance Levels of the listeners with minimal exposure to English and the native listeners. The Noise Tolerance Level was defined as the noise level at which the listener could repeat correctly about 50% of the sentences. Noise Tolerance Levels increased as the degree of familiarity and exposure to English increased. Buus and colleagues noted that the "...12 dB disadvantage is the same as that experienced by native listeners with a 60 dB hearing loss relative to normal listeners" (p. 897). In another study by Florentine (1985), 16 non-native speakers from a variety of languages were compared with 13 native speakers of English. The purpose of this study was to determine whether native speakers and non-native speakers take advantage of context to the same degree, and whether there was a difference between the two groups in the rate of improvement with decreasing noise levels. The non-native speakers were studying or teaching at the university level and were described as highly fluent speakers of English. The *SPIN* test, which consists of sentences varying in predictability presented in the presence of multi-talker babble noise, was used. Results indicated that non-native speakers had more difficulty understanding speech in the presence of noise than native speakers despite

their high level of fluency. Native speakers were able to achieve a 36% higher percentage of accuracy than non-native speakers on the high-predictability sentences and 27% higher on the low-predictability sentences. Florentine suggested that high-predictability sentences accentuate the difference between the native speakers and the non-native speakers because the native speakers gain more from predictability than non-native speakers. These findings indicate that "even highly proficient listeners...may lose as much as 30% of the information gathered by native listeners in marginal listening situations" (p. 1024). There is a point where speech is so degraded by noise that it is no longer intelligible with or without the benefit of contextual cues or influenced by degree of accent. Only speech that has a certain degree of overall intelligibility has the potential for further improvement with increased cues. Contextual cues, for example, may fail to upgrade the intelligibility of speech that is severely degraded (Sitler, Schiavetti, & Metz, 1983).

Mayo, Florentine, and Buus (1997) studied the effects of age of second language acquisition on the perception of non-native speakers of English. They compared a group of three listeners who were considered bilingual in Mexican-Spanish and English, nine native Mexican-Spanish-speaking listeners who learned fluent English before age 6, and a group of nine Mexican-Spanish-speaking listeners who learned fluent English after age 14 with nine monolingual English listeners. The *SPIN* test, which controls for linguistic predictability and is presented with a competing multi-talker background noise, was used. Listeners were asked to identify the final word of every sentence presented. Results of this study indicated that monolingual and early bilingual speakers

were less affected by noise and were able to demonstrate a greater benefit from context than those listeners who learned to speak English at a later age. Mayo and colleagues concluded that even highly fluent non-native listeners are affected by the age of acquisition for the efficient processing of a second language, especially in the presence of noise.

Studies of "filtered" speech have been conducted to determine how much of the speech signal is actually necessary for speech perception (Denes & Pinson, 1993). Such research was motivated by the need to determine the effect on intelligibility when speech is heard over transmission systems that respond only to a limited range of frequencies (i.e., telephones, hearing aids, and recording systems). Devices that respond only to certain frequencies are referred to as filters. Experimenters with this focus concluded that speech remains intelligible even if we hear only part of the speech spectrum. Pierce and David (1958, as cited in Bergman, 1980) determined that word intelligibility was equally affected at the critical frequency of 1800 Hz. That is, if the signal was low-passed at 1800 Hz or high-passed at 1800 Hz, the percent intelligibility score for words was about 67%. If, however, the filter has a total bandwidth of 1500 Hz encompassing the 1800 Hz frequency, reasonable conversational intelligibility can be expected in most instances. This agrees with Denes and Pinson's statement that a narrow band width of 1000 Hz in the range of 1500 Hz is sufficient to give a sentence articulation score of about 90%. Intelligibility increases as the bandwidth is broadened to include the frequencies between 100 Hz and 3000 Hz. Miller (1951, as cited in Bergman, 1980) found that with a bandwidth allowing frequencies above 300 Hz and

below 3000 Hz to pass, there would be little effect on the intelligibility of test words. Denes and Pinson concluded that there is nothing critical about a specific spectral area; if we discard it and listen to another part of the speech spectrum, we still get an intelligible signal. When listeners heard only those components of the speech wave below 2000 Hz, they could follow a conversation. However, they found that speech is equally intelligible if those low frequencies are eliminated, and only the components above 2000 Hz are heard. According to Bergman (1980), speech perception studies have provided clear evidence that high frequencies, which are especially important for the perception of consonants, are the main carriers of speech intelligibility.

Distortions of the speech signal may result from peak clipping or interruptions in the signal. Peak clipping occurs when incoming intensities exceed a predetermined output level of the transmissions system (amplifier). The perceived result of peak clipping may give speech a monotonous quality; however, physically, it results in severe waveform distortions. Severely distorted speech, distorted by the process of peak clipping, has been found to considerably alter speech quality; however, word articulation scores of 80-90% still can be obtained. Intelligibility is affected very little by such severe waveform distortions. Studies investigating interruptions in the speech signal have shown that if the signal is switched on and off at regular intervals, and the duration of each interruption is always equal to the duration of the speech signal allowed to pass (at one second intervals), whole words are lost and intelligibility is poor. When the rate of interruption increases to more than 10 interruptions per second, the word perception score rises to approximately 90%. This means that speech with periodic

interruptions at a rapid rate, interrupting as much as half the signal, will remain intelligible (Denes & Pinson, 1993).

The conclusion drawn from the combination of these studies is that the speech signal is "robust." That is, no one part of the speech wave is indispensable for satisfactory perception. The multiple acoustic cues available for perceiving speech reinforce one another. When one cue is eliminated, others remain. However, this conclusion may only apply to speech perception of a language in which one is fluent. When we listen to a foreign speaker, many of the acoustic cues deviate from the native listener's expectation, resulting in altered perception. Therefore, when multiple cues are distorted or eliminated by noise or interruptions in the signal and further complicated by an unfamiliar accent, the intelligibility of the signal can become significantly degraded.

<u>Multiple modalities</u>. Other nonlinguistic factors that appear to influence speech perception include the use of multiple modalities. Several investigators (e.g., MacDonald & McGurk, 1978; Massaro, 1987; McGurk & MacDonald, 1976; Miller & Nicely, 1955; and others as cited in Gagne, 1994) have demonstrated that normal-hearing adults are influenced by visual speech information. Even infants as young as 4 months of age can make use of visual information available in the speech message (e.g., Dodd, 1979; Kuhl & Meltzoff, 1982; Spelke, 1979 as cited in Gagne, 1994). The effect of visual information is often observed in communication exchanges between native speakers and non-native speakers. Accurate perception of the message appears to be facilitated by the physical presence of the speaker. Because telephone communication involves filtering as well as a loss of visual information, individuals often report more

difficulty understanding a non-native speaker over the telephone than in person. Visual information not only provides complimentary information regarding articulatory gestures, but non-verbal facial and body gestures also facilitate one's ability to perceive a message.

The idea that multiple modalities play a role in speech perception raises an important question of how to accommodate for this effect in research and clinical assessment. Tyler (1994) suggested that there is a need to increase our knowledge regarding the effects of auditory and visual noise on speech perception performance. Typically, investigations have been conducted in laboratory settings using a known, controlled and repeated stimulus in a quiet or noise "controlled" situation. In most cases these laboratory settings are not representative of typical listening situations that might affect speech perception performance. According to Gagne (1994), the audiology research involving speech perception has focused primarily on unisensory capabilities. This has been true of cross-linguistic research as well. With the exception of infant studies (e.g., Kuhl & Meltzoff, 1982; 1984; MacKain et al., 1983; as cited in Kuhl & Iverson, 1995), few references to the use of multiple modalities are to be found in cross-linguistic speech perception studies. Japanese researchers have taken the lead in this area. For example, Imaizumi (1997) reported a study involving the neural processes of audio and visual modalities. Findings suggested that visually presented articulatory information significantly affects L2 speech perception, and audio-visual training has the potential to build up proper neural representations of L2 phonetic categories. Akahane-Yamada and Tohkura (1997) reported a series of studies investigating the effects of

audio and audio-visual training of non-native listeners. Their results demonstrated that audio/visual training improved perception and facilitated integration of auditory and visual information. Training in perception transferred to improvement in production, and production training improved perception.

Measures of Speech Intelligibility

The concept of speech intelligibility usually implies a method of measurement in order to quantify an outcome. Speech intelligibility measures have been used for multiple purposes by different professions. Intelligibility measures were first used to evaluate the distortion of speech passed through different transmission systems, especially telephones (Fletcher, 1953). Communication engineers continue to use speech intelligibility tests while varying parameters, such as signal-to-noise ratio and bandwidth, to evaluate the effect of these parameters on the transmission system (Schiavetti, 1992). Audiologists also use speech intelligibility tests for similar purposes. Speech intelligibility tests are used to evaluate the quality of one transmission system (hearing aid) compared to another in determining the best system for the hearing impaired individual. In addition to the evaluation of hearing aid benefit, audiologists use speech intelligibility tests to evaluate the speech discrimination or recognition abilities of hearing impaired persons (Penrod, 1985). Linguists use speech intelligibility measures to determine whether two related speech varieties are to be considered as different dialects of the same language or as two different languages based on the mutual intelligibility of the two speech variations (Comrie, 1987). Speech-language pathologists have traditionally used intelligibility measurements as an index of severity

for speech disorders and to quantify improvement of these disorders. In recent years, linguists and speech-language pathologists have attempted to use intelligibility measures to quantify and assess the degree of foreign accentedness of non-native speakers of English.

Traditionally, measures of speech intelligibility have involved listener ratings because of their ease of administration. These methods typically include word identification tasks that require the listener to write down what the speaker says. The listener's identified words are compared with the speaker's intended words to determine a percentage of speech intelligibility. A second common method involving listener ratings is the use of a scale that allows the listener to make a judgment about the speaker's intelligibility. This type of procedure uses techniques such as equal-appearing interval scales, direct magnitude estimation, or intelligibility percentage estimates based on the listener's overall impression of the speaker. A third method of intelligibility rating involves acoustical measurements that attempt to correlate the physical parameters of speech with intelligibility.

Word identification tests

Stand-alone word identification tests have been used by communication engineers to measure speech intelligibility while evaluating the efficiency of speech transmission and by audiologists for the evaluation of speech recognition ability of the hearing impaired. Speech-language pathologists use either word identification or scaling procedures or both (Schiavetti, 1992). Word identification scores are derived from transcriptions and are usually calculated as a percentage of words correctly heard or as a

proportion that can be easily converted to a percentage. The chief advantage of word identification tests is that they produce a measure of speech intelligibility that is used easily by the researcher or clinician and is in a form that can be communicated with other professionals and laypersons. Research by Beukelman and Yorkston (1979) has shown a strong correlation between information transfer and word identification tests with dysarthric speakers. To measure information transfer, listeners were asked to answer 10 questions about the content of paragraphs recorded by dysarthric speakers. The same speakers were also measured on an isolated word test and a contextual speech intelligibility measure. A comparison of the three measures showed a strong correlation between the information transfer measure and both the isolated word and contextual speech intelligibility measures. These findings suggest a good criterion validity for the two word identification test intelligibility measures. Beukelman and Yorkston (1980) also found that word identification tests were more sensitive and accurate, especially in the midrange of intelligibility, than were scaled scores of passages. Scaled scores often overestimated intelligibility and, according to Samar and Metz (1988, as cited in Schiavetti, 1992), allowed for an unacceptably wide margin of error.

Another advantage of word identification tests is that they provide data for acoustical analysis. It is through acoustical analysis that speech characteristics can be examined in an attempt to correlate intelligibility with particular parameters of speech. Word identification tests can be specifically designed to contain distinct dimensions of speech for analysis of intelligibility such as voice onset time differences among voiced/voiceless consonant pairs (Samar & Metz, 1988 as cited in Schiavetti, 1992).

For example, word identification tests of speech intelligibility can provide more than degree of intelligibility. They also can provide useful data for explaining intelligibility deficits (e.g., Kent & Weismer, 1989; Kent, Weismer, Kent, & Rosenbeck, 1989; Weismer, Kent, Hodge, & Martin, 1988). Finally, analysis of word identification tests of speech intelligibility indicates that these measures are at least as reliable as those yielded by scaling procedures (Schiavetti, 1992). For example, when examining the speech intelligibility of dysarthric speakers, Yorkston and Beukelman (1978) reported that although intra- and inter-listener agreements were good for both word identification tests and scaling procedures, listener reliability was somewhat better for word identification tests. Others have also found the reliability of word identification tests to be high. For instance, Samar and Metz reported strong correlations when comparing the interscorer and intrascorer reliability of contextual word identification test results from speakers with hearing impairment. They reported a reliability of +.985 in both instances. Metz, Samar, Schiavetti, Sitler and Whitehead (1985) compared the reliability of isolated word and contextual word identification tests of the intelligibility of speakers with hearing-impairment. They reported the reliability of the sentences lists as +.95 and the reliability of the isolated word list as +.93.

Scaling procedures

Equal-appearing interval scaling. Equal-appearing interval scaling is the most common method of interval scaling used in the studies of speech intelligibility (Schiavetti, 1992). The listener assigns either a numerical rating or a descriptive rating. The National Technical Institute of the Deaf (NTID) developed an equal-appearing

interval scale for the purpose of measuring intelligibility. The NTID scale provides a scale of 5 possible descriptors of intelligibility. A rating of 1 is used to describe speech that is completely unintelligible, whereas a rating of 5 describes speech as completely intelligible. Listeners are presented with speech samples and asked to rate each sample based on the 5-point scale. For this particular scale, listeners are familiarized with the task by listening to a few examples that fall along the range from low to high intelligibility.

There is some disagreement as to whether the equal-appearing interval scale is appropriate for the measurement of intelligibility. According to Schiavetti (1992), the construct validity of interval scaling of speech intelligibility may be questioned. Construct validity is the degree to which a particular test or measuring instrument actually measures intelligibility (Maxwell & Satake, 1997). Stevens (1975) and Schiavetti, Metz, and Sitler (1981) described intelligibility as a “prothetic” dimension, reflecting variations in magnitude or quantity rather than quality. Schiavetti et al., used the term prothetic to characterize intelligibility because it is described by degrees along a continuum not easily partitioned into equal intervals. When intelligibility is measured in this way, it is based on an ordinal scale of measurement without the benefit of equal distances or differences on the scale. Stevens pointed out that if observers try to partition a prothetic scale into equal intervals, they typically demonstrate a systematic bias by subdividing the lower end of the continuum into smaller intervals than the upper end of the continuum. This inequality of intervals along a prothetic continuum may be the result of variations in the abilities of listeners to discriminate along the continuum.

A metathetic continuum, in contrast, is described as one in which the dimension varies in a qualitative sense. Pitch is an example of a dimension that can be measured on a metathetic scale. For instance, it varies from high to low; it varies in terms of a change in quality. Measurement of a metathetic, qualitative dimension is considered to be an appropriate use of an equal-interval scaling procedure and will, therefore, offer greater reliability in measurement. Because intelligibility is considered by some (e.g., Schiavetti et al., 1981; Stevens, 1975) to be prothetic, direct magnitude estimates or word identification tests are considered to have better construct validity than equal-interval scales.

Direct magnitude estimates. Direct magnitude estimates allow each listener to judge each speech sample with a number that is proportional to the perceived ratio of speech intelligibility. Listeners may or may not be provided with a standard reference. When used, a standard reference will usually represent the lower, middle, or upper portion of the intelligibility range to "calibrate" the listener. Listeners then assign a number to the subsequent speech samples that they feel represents the degree of intelligibility of the speaker. When the listener is not provided with a standard reference, he/she assigns any number to the first speech sample and then assigns numbers to subsequent speech samples accordingly. These numbers correspond to the ratios of the perceived magnitudes of the intelligibility of the various speech samples. Unlike interval scaling, direct magnitude estimates are less constrained to fit ratings into a defined linear scale (Schiavetti, 1992). Schiavetti stated that if a scaling procedure is necessary, direct magnitude estimation is a viable scaling procedure. However, direct

magnitude estimation may not be the most practical method for measuring speech intelligibility. For example, the result of direct magnitude estimation is a scaled value without a common unit of measure such as percentage of words heard correctly. This makes interpretation and communication of the data to other professionals or laypersons more difficult. Direct magnitude estimation also is used best for the measurement of a large number of stimulus samples along the dimension to be scaled. Its use is contraindicated when only one or relatively few samples are measured in a single instance. Comparisons of reliability between scaled measures and word identification tests have resulted in good intra- and inter-listener agreement; however, word identification tests have a somewhat better consistency than scaling procedures (Yorkston & Beukelman, 1978).

Intelligibility percentage estimates. Another method of scaled measurement is to ask listeners to assign a percentage intelligibility score to a speaker after listening to a paragraph-sized sample of speech. The percent intelligibility estimate is based on the listeners' impressions of how well they could understand the speaker from the overall reading of a passage. Beukelman and Yorkston (1980) found that both trained and naive listeners who estimated the intelligibility of dysarthric speakers by assigning a percent intelligibility score after listening to the reading of a passage consistently overestimated the intelligibility of the speakers when compared to word identification scores derived from transcriptions. They also noted wide variability among judges, especially in the moderate and severe speech samples. Further investigation of the data suggested that as listeners became familiar with the passage, the intelligibility percentage estimates

increased. This observation might explain the discrepancy between intelligibility percentage estimates and actual word identification scores.

Acoustical measurements

Acoustical measurements have been helpful in discovering acoustic or phonetic contrasts that are specific to a communication impairment and that may contribute to speech intelligibility. For example, Weismer and Martin (1992) reported that the mean slope of the second-formant (F2) transitions are highly correlated with the word-recognition intelligibility scores of dysarthric individuals with amyotrophic lateral sclerosis. Acoustical measurements have also been found to provide a reasonable prediction of intelligibility of hearing impaired speakers based on parameters such as consonant voicing contrasts (Metz et al., 1985; Monsen, 1978). The richness of the acoustic-perceptual signal is sometimes seen as a disadvantage due to the overwhelming number of measures it provides. Research on speech perception, voice quality, acoustic phonetics, and speech disorders has helped to limit the selection of acoustic measures that are most highly correlated with intelligibility (Kent, 1992). Further acoustical research may establish how the acoustical output of a non-native speaker differs systematically from that of native speakers of English. For example, Arslan and Hansen (1997) studied temporal features and frequency characteristics in English produced by native Mandarin, German, Turkish, and American English speakers. They were able to identify acoustical differences between foreign speakers that differentiated native speakers of one language from those of another. Such information should be helpful in preparing effective materials for improving the intelligibility of non-native speakers as

well as in establishing explanations as to why native speakers of English have difficulty understanding non-native speakers.

Probably no measure of speech intelligibility can be seen as "best." Kent (1992) concluded that "it may be that no single test will ever satisfy research and clinical needs...the study of intelligibility might be undertaken with several tools, including word-intelligibility tests, sentence-intelligibility tests, rating scales and others as appropriate" (p. 8).

Intelligibility Measurements Used Specifically for Non-Native Speakers of English

A number of speech intelligibility measures for the purposes of assessment and research with non-native speakers of English have been devised and many are currently in use, although there is still no single preferred method of assessment. Lane (1963) measured intelligibility by counting the total number of words listeners transcribed correctly; Barefoot, Bochner, Johnson, and VonEigen (1993) counted percentages of key words recognized; Brodkey (1972) analyzed the accuracy of paraphrases; and Fayer and Krasinski (1987) and Palmer (1976) asked listeners to rate intelligibility directly on a Likert scale. Gass and Varonis (1984) asked listeners to write out sentences produced by non-native speakers and then assigned scores based on deviations between the transcripts and the intended utterances. Munro and Derwing (1995a) asked listeners to do a similar transcription as well as to assign a perceived comprehensibility judgment using a 9-point Likert scale. They examined the relationships between scores of comprehensibility and direct transcription with global foreign accent scores. Morley (1993) developed a Speech Intelligibility Index that consists of listener ratings of a tape

recorded sample of impromptu speech. Listeners provide two ratings of the sample on a 6-point scale based on a description of the intelligibility of the individual's speech and the interference of accent with communication.

On a broader scale of language proficiency, Hinofotis and Bailey (1980) developed an Oral Communication Rating Instrument used to assess videotaped speech samples for the purposes of rating oral communication of foreign students. It consists of three main sections that include: initial overall impression, performance categories, and final overall impression, each of which is based on a 9-point Likert scale. Pronunciation is only 1 of 12 subcategories of performance ranked by listeners in degree of importance. *The Test of Spoken English* (*TSE*) was developed to provide a standardized measure of oral language proficiency (Clark & Swinton, 1979, as cited in Stansfield and Ballard, 1984). It is administered and scored by the Educational Testing Service (ETS) for the purpose of measuring oral English proficiency. Tape-recorded samples are taken from the speaker and sent to ETS for scoring. Raters are trained at one-day workshops and are experienced teachers and specialists in the field of English as a second language. Each sample is rated independently by two raters, and the examinee's score is an average of the two ratings. Scores are assigned for overall comprehensibility, pronunciation, grammar, and fluency. Retired *TSE* test forms, referred to as *Speaking Proficiency English Assessment Kits* (*SPEAK*), are used by colleges and universities to assess the spoken English skills of foreign teaching assistants and other foreign students. They are also used within the health-related professions, government agencies, and private corporations (Stansfield & Ballard, 1984).

A Specific Test of Intelligibility: *The Speech Perception in Noise* Test

It has been noted by Schiavetti, Sitler, Metz and Houde (1984) that contextual speech intelligibility tests have more external validity as measures of real world speech intelligibility than do isolated word intelligibility tests. The *Speech Perception in Noise* (*SPIN*) is an example of a speech intelligibility test that relies on direct transcription but also takes context into account. The *SPIN* test was developed by Kalikow et al. (1977) as a speech recognition test designed to assess a listener's ability to use linguistic-situational information in speech in contrast to the use of acoustic-phonetic information only. The test is designed to represent everyday listening situations in which noise interferes with the understanding of speech. The *SPIN* sentences were recorded with the background noise produced by several speakers, referred to as multi-talker babble. Multi-talker babble noise masks some of the sounds, so that the listener has less acoustic information on which to base the interpretation of the acoustic signal. The listener hears a recording of a list of sentences presented with a background of multi-talker babble and repeats or writes the last monosyllable (target word) of each sentence. Each of the 50-sentence lists contains 25 sentences in which the target word is related to the context of the sentence and that are referred to as high-predictability (HP) items. For example, "The watchdog gave a warning *growl*" is a high-predictability item. The remaining 25 target words were designed to be primarily identified through acoustic-phonetic cues (contextually neutral). They occur in sentences that offer minimal contextual cues and are considered to have low-predictability (LP) items (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984; Morgan, Kamm, & Velde, 1981). "Mr.

Smith thinks about the *cap*" is an example of a low predictability item. Morgan et al. (1981) conducted a form equivalence study of the 10 forms of the *SPIN* and found 7 of the 10 lists to be equivalent forms. The *SPIN* test was originally developed with consideration of the familiarity of the final (key) words. Key words were eliminated if they were considered to be too seldom used or very frequently used words in English. Key words with frequency counts in the range of 5 to 150 per million words were chosen from the Thorndike-Lorge list (as cited in Kalikow et al., 1977).

The original purpose of the *SPIN* test was to assess the speech perception of hearing-impaired individuals in the presence of different levels of noise in a speech context representative of the individual's day-to-day situations. It has been used in several clinical studies and with a variety of different clinical populations (Dirks, Kamm, Dubno, & Velde, 1981; Elliott, 1979; Hutcherson, Dirks, & Morgan, 1979; Owen, 1981). However, Bilger et al. (1984) reported that the *SPIN* test has been used primarily on young, normal-hearing adults for statistical evaluation and standardization.

The *SPIN* test has been used in the study of non-native listeners' perception of American English in noise (e.g., Florentine, 1985; Florentine, Buus, Scharf, & Canevet, 1984; Mayo et al., 1997). For example, Florentine (1985) investigated the ability of non-native listeners and native listeners to take advantage of linguistic context in the presence of babble noise. The conclusion of this study supported others (Bergman, 1980; Florentine et al., 1984; Nablelek & Donahue, 1984) who have found that non-native speakers may demonstrate native-like speech recognition in quiet but have more difficulty understanding speech than native listeners in the presence of background

noise. Florentine also concluded that in the presence of noise, non-native listeners did not benefit as much from contextual cues as did native listeners.

The SPIN sentences have also been used to investigate the processing effort associated with recognizing and comprehending accented speech in comparison to native-sounding speech. Schmid and Yeni-Komshian (1999) used the SPIN sentences produced by native speakers and non-native speakers of English. In this study, the SPIN sentences were modified by placing intended mispronunciations in the target words. The listeners were asked to identify mispronunciations as soon as they were identified. The listeners response times were measured for comparison. Schmid and Yeni-Komshian found that listeners were able to detect more mispronunciations in the speech of native speakers than non-native speakers. They concluded that although non-native speakers are judged as intelligible, listeners may have to expend more effort to recognize and comprehend accented speech in comparison to native-sounding speech.

Problems with Assessing Speech Intelligibility

"Intelligibility is considered the most practical single index to apply in assessing competence in oral communication" (Subtelny, 1977, p. 183) of individuals with foreign accents or distorted speech resulting from a communication disorder. Although the importance of speech intelligibility has been clearly identified, there is no consensus as to the best methods of measurement and assessment (Munro & Derwing, 1995a). The task of measurement and assessment has been complicated by numerous factors that involve both the speaker and the listener.

The first problem with the assessment of speech intelligibility is that intelligibility does not rise entirely from the sounds produced by a speaker. The burden of any communication interaction is shared equally between the speaker and the listener, and both parties have a significant impact on the success of communication. Variability among listeners in their tolerance levels, experience with listening to non-native speakers, and topic familiarity is as great as the variability in the accent of the speaker. Both speaker and listener play a role in the process of judging intelligibility. According to Schiavetti (1992),

> it is crucial to understand that any measure of speech intelligibility is a measurement of the interaction between a speaker, a transmission system, and a listener. Therefore, it is important to quantify the parameters that concern the speaker's production, the quality of the transmission system, and the listener's response. (p. 12)

Auditory-perceptual measures are common throughout the literature addressing speech intelligibility (e.g., Anderson-Hsieh et al., 1992; Flege, 1992; Munro & Derwing, 1995a). The benefits of auditory-perceptual measures include convenience, economy, and usefulness for assessment of treatment outcome (Kent, 1996; Metz, Schiavetti, & Sitler, 1980). Auditory-perceptual measures also take advantage of the ability of the auditory system to understand speech under various conditions. However, the weaknesses of auditory-perceptual judgments are numerous. One problem with the assessment of intelligibility is that although two listeners may share a common idea of what intelligibility is, they may use very different methods to measure it and to understand its correlates in the act of speaking (Kent, 1992). Even though listeners may disagree in their ratings of a message, they often cannot perceptually identify specific

components affecting their judgment because of the perceptual tendency to hear all aspects of the signal as a whole (Kent, 1996).

A weakness of auditory-perceptual measurements is that judges may not agree on the definitions for the perceptual dimensions to be rated (i.e., pronunciation, intelligibility, comprehensibility, interpretability) or even which dimensions should be rated (Kent, 1996). The perceptual dimensions to be rated may vary from one investigation to another or from one listener to another. For example, there is still no agreement as to which aspects of pronunciation are most crucial for intelligibility. Munro and Derwing (1995a) cited several studies (Albrechtsen et al., 1980; Gimson, 1970; Johansson, 1978; Schairer, 1992) that have attempted to establish a hierarchy of pronunciation errors. Differences in target languages and research methodologies have complicated the conclusions. Gimson, for example, argued that native-like production of consonants is more important than the production of vowels in the comprehension of English. In contrast, Schairer came to the opposite conclusion when studying native English speakers learning Spanish. Other researchers such as Anderson-Hsieh et al. (1992), Johansson (1978), and Palmer (1976) concluded that prosodic errors were more detrimental to comprehension than segmental errors.

From the above, it appears that specialists often fail to reach consensus on which perceptual dimensions should be rated to assess intelligibility (e.g., Fayer & Krasinski, 1987; Munro & Derwing, 1995a). Intelligibility could be evaluated on a multidimensional scale that might include prosody, grammar, and phonetic substitutions. Variations in the definition of prosody may confuse listeners. For example, some

listeners may define prosody simply as the parameters of speech that are perceived as pitch, intensity, and duration (Kent & Read, 1992). Anderson-Hsieh et al. (1992) described the features of prosody as timing, rhythm, intonation, and stress. Others may include pausing and speech rate in their definition of prosody (Kent & Read, 1992). It is important to control for this variability in definitions and parameters to be assessed. Definitions should be provided and training with reference samples should be used to promote better interjudge agreement (Kent, 1996).

The proficiency of the auditory system provides the normal listener with the ability to make sense out of the total signal. This explains our ability to understand speech even under degraded conditions (Kent, 1996). This proficiency also can be considered a weakness in studies involving auditory-perceptual assessment because the human auditory system is limited in its ability to rate various perceptual dimensions independently. In other words, one dimension is often influenced by co-occurring dimensions. Listeners can often determine that there is something "unusual" about the speech signal yet be unable to describe exactly how or why the message is distorted (Orlikoff & Baken, 1993).

It would seem that use of acoustic analyses should provide a means for avoiding these problems with perceptual judgments; however, acoustical analyses are not without problems. Instrumental (acoustic or physiological) analysis would likely provide greater reliability and possibly improved accuracy over auditory-perceptual measures alone. Unfortunately, many studies have shown only weak associations between acoustic measures and perceptual ratings. For example, Arends, Povel, Van Os, and Speth

(1990) reported discouraging results with such a correlation for the speech of the deaf. Kent et al. (1994) also found a poor association between acoustic measures and perceptual measures in the judgment of dysarthric speakers. Kreiman, Gerratt, Precoda, and Berke (1992) observed that expert judges differed considerably with their perceptual ratings of pathological voices when compared to acoustic measures. They concluded that a fixed set of acoustic measures may not correlate highly with perceived severity across a range of abnormal voices across different judges.

Researchers (e.g., Metz, Samar, Schiavetti, & Sitler, 1990; Weismer, Kent, Hodge & Martin, 1988; Weismer & Martin, 1992) have attempted to identify an ideal acoustical model that would relate intelligibility to the physical properties of speech and would, in turn, relate these properties to the movements and activities involved in the speech production process (Nickerson & Stevens, 1980). However, this ideal model of speech intelligibility is nearly impossible to identify due to the complexity and variability inherent in normal speech. For example, different productions of a given utterance may vary considerably with respect to objectively measurable properties while remaining highly intelligible. What constitutes a normal range of values for many of the measurements that could be obtained on a single utterance would likely depend to some degree on linguistic and situational context. Finally, the interpretation of measurements often will be interdependent. The values obtained for any one dimension may be influenced by co-occurring dimensions. Nickerson and Stevens (1980) pointed out that "intelligibility may prove to be sensitive to the interactions among various objective properties as well as to their combined individual effects" (p. 339). For

example, as one property changes, others are likely to be affected. If intelligibility is improved as a consequence of training, it is impossible to conclude that the improvement was a direct result of the modification of the properties on which training was focused unless there is certainty that none of the other properties of speech were changed as well. On the other hand, if there is no improvement as the result of training, it cannot be concluded that the property in question is not important to intelligibility. It is possible that the changing of one component may not be enough to improve intelligibility; however, changing it may be a necessary step in the process that may not be immediately observable (Nickerson & Stevens, 1980). Osberger (1978, as cited in Nickerson & Stevens, 1980) studied changes in pause duration in deaf children's speech and found a decrease in intelligibility. Changes in the relative durations of stressed and unstressed vowels increased intelligibility by a small amount. The result of the Osberger study demonstrates that making speech more "normal" with respect to a particular feature (i.e., pause duration) does not necessarily increase overall intelligibility. This does not mean that pause duration has no relevance to intelligibility. The effect of pause duration may interact with that of others in complicated ways. In this example Osberger concluded that it may be that gross deficiencies in pause duration and durations of stressed and unstressed vowels may be sufficient to assure low intelligibility, whereas, proper timing is not sufficient to assure high intelligibility if the speech is deficient in other ways, such as with multiple segmental errors.

Accent and Intelligibility

Foreign accent has been defined as the consequence of a speaker's applying the phonological rules of a language, usually his/her first language, to the target language, instead of learning and applying new phonological rules (Wingstedt & Schulman, 1987). Thus, foreign accent reflects a difference in the pronunciation patterns of non-native speakers due to his or her first language backgrounds (Arslan & Hansen, 1997). Pronunciation patterns are driven by phonological rules that dictate which syllables are allowed to have stress, the language's phonemic inventory, and the constraints on sound combinations. Foreign accent is characterized by the speaker's transferring the known rules of a native language to productions in another language. Listeners perceive foreign accent as deviations in the expected pronunciation of their language. Major (1987) characterized foreign accent as a type of noise in the speech signal that can interfere with the message. A strong accent can cause the listener to strain to decipher the meaning of the message.

Several studies have shown that native listeners tend to view non-native speakers negatively simply because of their foreign accent (e.g., Anisfeld, Bogo, & Lambert, 1962; Brennan & Brennan, 1981, Kalin & Rayko, 1978; Lambert, Hodgson, Gardner, & Fillenbaum, 1960; Ryan & Carranza, 1975; as cited in Munro & Derwing, 1995a). As a result of the negative impact of a foreign accent, many programs offering second language instruction have focused attention on accent reduction without regard to specific features that may interfere with intelligibility. Munro and Derwing (1995a) stated that a reduction of accent does not necessarily result in an increase in

intelligibility. "Not only is there little empirical evidence regarding the role of pronunciation in determining intelligibility, but also there is no clear indication as to which specific aspects of pronunciation are most crucial for intelligibility" (p. 76).

In an attempt to investigate the relationship between foreign accent and intelligibility, Munro and Derwing (1995a) studied native listeners of English who were asked to rate comprehensibility and foreign accent from speech samples of native Mandarin speakers. They found that listeners tended to rate accent more harshly than they rated comprehensibility. In comparison, the accent scores were a much poorer reflection of the listeners' actual comprehension of an utterance than were the perceived comprehensibility scores. Listeners sometimes rated utterances as moderately or heavily accented even when they were able to transcribe them perfectly, indicating that the presence of a strong foreign accent does not necessarily result in reduced intelligibility or comprehensibility. The researchers concluded that foreign accent ratings did not predict intelligibility very well and inferred that, when judging accentedness, listeners may be influenced by variables that ultimately had no impact on whether or not the message was understood. For example, Munro and Derwing compared assessments of phonemic errors, phonetic errors, and goodness of intonation to listener ratings of accentedness, comprehensibility, and intelligibility. They found significant correlations between accentedness and these assessments. However, perceived comprehensibility and intelligibility showed weak correlations when compared to these measures. Nonphonological influences such as grammatical errors have also been shown to negatively influence pronunciation judgments (Munro & Derwing, 1995a; Varonis &

Gass, 1982). In summary, Munro and Derwing (1995a) stated that "the role of comprehensibility in accent judgments varies from listener to listener and that accent scores cannot be relied upon as a means of assessing comprehensibility. Moreover, accent scores are poorer indicators of intelligibility than were perceived comprehensibility scores" (p. 92).

Pronunciation

According to Anderson-Hsieh et al. (1992), errors in pronunciation fall into four general types: segmentals, suprasegmentals (prosody), syllable structure, and voice quality. The segmental component involves errors in consonants and vowels. Suprasegmental errors may involve timing, rhythm, phrasing, intonation, and stress. Errors in syllable structure include adding or deleting a segment or syllable. The component of voice quality refers to characteristics of pronunciation that affect entire utterances (Abercrombie, 1967: Laver, 1980 as cited in Anderson-Hsieh et. al., 1992). Although voice quality is a component of pronunciation, it has not been well investigated in second language learners. Examples of typical voice quality characteristics include a tendency to keep the lips rounded throughout speech, a tendency to keep the body of the tongue slightly retracted into the pharynx while speaking, or a persistent use of a whispered quality of speech (Laver, 1980 as cited in Esling & Wong, 1983). Esling and Wong (1983) stated that voice quality in non-native speakers may deviate from native speakers as a result of inappropriate posturing of the articulators, such as the tight-jawed posture and dentalized tongue body setting often found in Chinese ESL learners.

Segmentals. Of the four types of errors described by Anderson-Hsieh et al. (1992) segmental errors are the most salient, especially to naive listeners. Phonemic errors are often blamed by listeners who are untrained and unable to explain what they are hearing in any other way. This observation has been reported in studies involving undergraduate students rating international teaching assistants (e.g., Hinofotis & Bailey, 1980). However, although these errors are the most salient, researchers have found that even if the individual sounds and words of a language are pronounced correctly, a foreign accent will still be evident because of the transfer of intonation patterns of the native language to the target language (Chun, 1989; Munro, 1995).

Suprasegmentals. Suprasegmental errors (prosody) were defined by Anderson-Hsieh and her colleagues as errors in timing, rhythm, intonation, and stress. Intonation is the pattern of pitch variation across a word or group of words. It is important for carrying meaning as well as emotion and attitude. Rhythm is a timing mechanism. Languages are either syllable-timed or stress-timed. That is, the stress is applied in a systematic way that gives the language its characteristic rhythmic quality. For example, Japanese and Spanish are syllable-timed, whereas English and German are stress-timed. Stress occurs at regular intervals of syllables in syllable-timed languages and at regular intervals of time in stress-timed languages. Duration of stressed syllables in stress-timed languages is long, 1.7:1, and in syllable-timed languages it is much shorter, 1.1:1 (Celce-Murcia, Brinton, & Goodwin, 1996; Major, 1981).

According to Celce-Murcia et al. (1996), rhythm is a function of the number of syllables in a given phrase in syllable-timed languages. Phrases with an equal number of

syllables take roughly the same time to say. In stress-timed languages, rhythm is a function of the number of stresses in a phrase. In English the length of an utterance depends on the number of stressed syllables it contains; in a syllable-timed language, length of an utterance depends on the number of syllables.

Suprasegmental features are those features that are part of an utterance but are larger than segments. In other words, these features may involve several segments. Pitch rise, for example, may affect a whole syllable, a word, or even a phrase. Prosody varies across languages just as segmental cues do, and unusual prosodic patterns can interfere with the listener's ability to comprehend the message. As Nash (1972) phrased it, appropriate prosody is like good background music in a movie. If it is appropriate, we are hardly aware of it, but if it is not appropriate, the conscious mind must deal with it. According to Nash, an appropriate prosodic pattern enhances and confirms the lexical message, but an inappropriate prosodic pattern can deny or contradict the intended message. When this happens in a non-native language, the native speaker is not able to use the linguistic context because the preceding utterances also had an inappropriate prosodic pattern. The listener is unable to relate the total meaning of one utterance to the total meaning of the next. Inappropriate prosody then, seems to have a cumulative effect, sometimes frustrating the listener to the point of no longer wanting to put the effort into understanding the message. The bottom line is that suprasegmentals are important to draw the listeners attention to important information in the discourse. Second language learners are often so focused on learning the lexicon that they miss the overriding melody and rhythm of utterances (Anderson-Hsieh, 1992).

Some researchers have suggested that prosody is the factor that has the greatest impact on the listening comprehension of native speakers of English exposed to non-native speakers. For example, in a study comparing non-native speakers' errors in prosody, segments, and syllable structure with accent comprehensibility ratings, Anderson-Hsieh et al. (1992) found that listeners rated prosody as affecting accentedness and perceived comprehensibility to a greater extent than the other factors. Tajima, Port, and Dalby (1994) manipulated temporal variables in the speech samples of Taiwanese and Mandarin speakers to resemble native speaker patterns. They also manipulated the speech sample of a native English speaker to resemble the temporal pattern of a Chinese non-native speaker of English. The intelligibility of the non-native speaker improved significantly as a result of the adjustment in the temporal pattern. In addition, manipulation of the native speaker's utterances resulted in a reduction in overall intelligibility.

Syllable structure. Anderson-Hsieh et al. (1992) described syllable structure errors as the addition or deletion of a segment or syllable. Consonant deletion and vowel insertion are the most common types of errors. Many of the observable characteristics of foreign accent are due to the application of phonological rules of another language in speaking the second language (L2). Such rules may affect syllable structure as well as phonemes. For example, some languages such as Mandarin and Brazilian Portuguese do not have stops in the word-final position, whereas other languages such as Turkish have only voiceless stops in the word final position. An inappropriate carryover of such phonological patterns may result in inappropriate

syllable duration, inappropriate voicing, or the addition of a vowel. Arslan and Hansen (1997) studied acoustic differences among Turkish, Mandarin, German, and American English. They compared voice onset times, word final stop duration, durational parameters at the segmental and word level, the slope of the intonation contour as well as frequency analysis among the four languages. They concluded that, in general, the non-native speakers had longer word final stop closure duration than native English speakers. They also found this particular aspect to be the best indicator of foreign accent. Another example of the influence of phonological rules from the first language may be observed in listening to native speakers of Spanish when speaking English. In this situation, we often hear a case of "epenthesis," the addition of a segment. For example, it is quite common to hear non-native speakers adding an initial "e" to words that begin with /s/, "especial" for "special." This occurs because the phonological rules of Spanish do not allow words to begin with /s/ clusters. In Brazilian Portuguese there is a phonological rule which states that the syllable final /s/ becomes voiced when it is immediately followed by a voiced sound. Due to the carryover of this rule into English, "Yes I am" becomes "Yez I am." Another example of transferring a phonological rule from Brazilian Portuguese would include epenthesis after final stops. Phonologically, there are no final stops in the language. As a result, a word in English that ends with a final stop (i.e., *cat*) will result in the addition (epenthesis) of a reduced, very short vowel. In contrast, when words end with an unstressed but long /i/ (i.e., *happy*), the result will be a reduction of the vowel to a shorter and weaker version than the native English production.

Factors Influencing Accent

It should be clear from the previous section that many factors contribute to the perceived degree of foreign accent. These include the number and severity of segmental errors (Flege & Eefting, 1987; Gatbonton, 1975; Major, 1987; Ryan, Carranza & Moffie 1975 as cited in Flege, 1988), inappropriate use of stress, rhythm and intonation (Bond & Fokes, 1985; Fokes & Bond, 1984; Varonis & Gass, 1982; Willems, 1982 as cited in Flege, 1988), and a carryover of phonological rules from the first language that results in syllable structure errors (Anderson, 1983; Broselow, 1983, 1984; Karimi, 1987; Sato, 1984; Tarone, 1980 as cited in Anderson-Hsieh, et al., 1992). Research investigating factors that influence non-native speaker pronunciation in native Italian adults (Flege, Munro, & MacKay, 1995) found that age of learning, speaker's gender, relative use of the first and second languages, and length of residence also affected the degree of perceived accent.

Age of learning accounted for more variance than did any other factor. Listeners identified 78% of the native Italian speakers who began learning English before the age of 4 as authentic speakers of English. As age of learning increased, the number of speakers identified with a foreign accent increased. For example, of those who began learning English after the age of 12 years, only 6% were identified as meeting the criterion for authentic pronunciation. None of the native Italian speakers who began learning English after the age of 16 years met the criterion for authentic pronunciation. The effect of gender had a variable effect on pronunciation in relation to age of learning. Female speakers who began speaking English at an average age of 9.6 years were found to

pronounce English somewhat better than did males matched for age of learning. The male native Italian speakers who began learning English in late adolescence were found to pronounce English better than their age-matched cohorts. The amount of language use also explained 15% of the variance in foreign accent ratings. That is, the combination of language use for home, social, and work use is a factor influencing the degree of perceived foreign accent. Although length of residence has been identified as a factor, it contributed to less than 2% of the variance observed in a step-wise multiple regression analysis.

The literature suggests that there is some limitation in the ability of adults to speak a foreign language without an accent. The earlier in childhood a second language is learned, the greater the likelihood of acquiring a more native sounding accent (Kent, 1997). A great deal of research has been done on identifying the age after which one can no longer achieve a native-sounding accent in the L2. This is called the "critical age hypothesis." Patkowski (1994) along with others (e.g., Lenneberg, 1967; Scovel, 1969 as cited in Flege, 1988) stated that puberty (ages 12-15) is the cut off for speaking an L2 without accent. However, Flege and other researchers have found that children who began learning the L2 as young as 7.5 years of age can still have a detectable accent. Flege concluded that the critical age lies somewhere between the ages of 5-8. Several researchers have studied other factors that affect accent. For example, Purcell and Suter (1980) and Thompson (1991) found that aptitude for oral mimicry or self-ratings of oral mimicry influenced accent, along with other factors such as length of time in the L2 environment, age of arrival into the L2 country, and strength of concern for

pronunciation accuracy. Flege and Fletcher (1992) conducted a multiple regression analysis and found that the age of arrival into the L2 speaking environment (country/USA) accounted for 79% of the variance in accent scores. Formal English language instruction increased the R-square value to 85%, and none of the other variables, such as percentage daily use of English, gender, and chronological age, were found to correlate with degree of foreign accent.

From another perspective, studies have focused on factors that affect the comprehension of accented speech. Varonis and Gass (1982) concluded that grammar and pronunciation interact to influence the overall intelligibility of non-native speakers. In a later study Gass and Varonis (1984) identified familiarity issues as an additional factor in comprehending non-native speakers. They concluded that familiarity with the topic, familiarity with non-native speakers in general, and familiarity with a particular accent and a particular speaker all had an effect on intelligibility. Anderson-Hsieh et al. (1992) identified prosody as affecting accentedness and perceived comprehensibility; however, they did not measure intelligibility or determine the relationship between prosody and intelligibility.

<u>Foreign Accent Ratings</u>. Major (1987) coined the term "global foreign accent," which he defined as "...overall pronunciation proficiency in a second language, or how native-like the accent is" (p. 157). Foreign accent ratings are typically based on the judgments of native listeners, usually by means of scaling procedures. For example, Varonis and Gass (1982) asked listeners to judge accent on a 5-point scale. They used expert raters to determine an absolute scale that was then compared to the ratings of

naive listeners. Fayer and Krasinski (1987) also had listeners judge pronunciation and intonation on a 5-point scale. Munro and Derwing (1995a) asked listeners to rate the degree of foreign accent on a 9-point scale where 1 represented no foreign accent 9 represented a very strong accent. Munro and Derwing found that accent ratings were significantly correlated with all the error types they studied, such as phonetic, phonemic, and grammatical errors and goodness of intonation ratings, but accent ratings were not correlated with intelligibility ratings. Hinofitis and Bailey (1980) did not directly assess "foreign accent" but had listeners rate overall language proficiency, of which pronunciation was a component, on a 9-point Likert scale.

Scales that include magnitude estimation techniques have been described as better suited to a listener's ability to resolve differences in degree of foreign accent than other methods, such as equal-interval scaling (Major, 1987). The magnitude estimation technique has been used by several researchers (i.e., Flege, 1988; Flege & Fletcher, 1992; Flege, Munro & MacKay, 1995; Major, 1987). This procedure requires listeners to estimate degree of foreign accent by moving a lever on a response box after hearing the stimulus. The range of lever movement would have been previously defined by the labels "no foreign accent" at the top, "medium foreign accent" at the middle, and "strong foreign accent" at the bottom of the scale. No visible number scale would be available to the listeners. The lever typically is attached to a potentiometer that yields a score ranging from 1 to 256. In the studies mentioned no foreign accent at all was represented by a score of 256, whereas a rating of 1 represented the strongest possible foreign accent. Once the lever was placed by the listener, the listener pressed a button that then

calculated a score based on this scale of 256. Flege and Fletcher (1992) stated that this method represents a fine scale that is more sensitive to perceptible judgments than a courser 7-point scale might be, and that it can measure more subtle perceptual differences.

Country of Origin. It seems logical that the more different the native language is in phonological structure, phonetic inventory, and prosodic differences, the more influence these characteristics would have on the intelligibility of the non-native speaker. However, very little research has been done to compare the intelligibility of non-native speakers of different language backgrounds. Country of origin is rarely identified in the literature as an important predictor of intelligibility. As mentioned earlier, Flege and Fletcher (1992) found that age of arrival into the L2 speaking community and the number of years of formal English language instruction were the most significant predictors of the perceived degree of foreign accent. In an attempt to develop a profile that would predict the non-native speakers who were most likely to pronounce English well, Purcell and Suter (1980)

identified country of origin as a factor. They compared four languages--Arabic, Persian, Japanese, and Thai--and found that the Arabic and Persian speakers in their study were favored over the Japanese and Thai speakers. This article did not provide methodological descriptions for how many speakers were included and how the researchers arrived at this conclusion. Purcell and Suter did conclude that non-native speakers who were good mimics, lived in an English speaking country for a number of years, and for most or all of that time have resided with a native speaker of English; and

were concerned about the accuracy of their English pronunciation would represent the speakers with the best pronunciation. Gender was not found to have a significant influence on pronunciation.

Gallego (1990) compared three non-native speakers of English, one speaker from Korea, one from Italy, and the third speaker of Hindi from India. The speakers were judged by native speakers of English who found the native Italian speaker to be the easiest to understand, the native Korean speaker the most difficult, and the Hindi speaker to be somewhere in between. This judgment was based on the total number of communication breakdowns identified per speaker. However, when the communication breakdowns were calculated per 100 words, the native Korean speaker had almost three times as many communication breakdowns as the Italian and Hindi speakers.

Hinofotis and Bailey (1980) compared the ratings of 10 undergraduate students with a group of three experienced English as a second language teachers and three instructors responsible for training foreign teaching assistants. They asked the raters to respond to a questionnaire rating areas of non-native speaker communication that may be problematic. The questionnaire included 20 statements that were rated on a 9-point Likert scale to indicate the raters degree of agreement or disagreement. Two of those items were used to determine whether the raters felt that native Oriental speakers of English were harder to understand than Europeans speaking English. Although both groups of raters were found to be in agreement that Oriental speakers were harder to understand, the undergraduate responses were somewhat stronger than the other group.

Hinofotis and Bailey cautioned that these findings are based on what raters *think* versus whether Oriental speakers are actually harder to understand.

Statement of the Problem

The Need for this Research

Speech intelligibility is a topic that has fascinated researchers trying to determine what makes one speaker more difficult to understand than others. The review of the literature presented above provides examples from the large body of literature devoted to speech intelligibility issues in general in addition to a review of research specific to the area of non-native speakers of English. It is clear that communication is often complicated by a non-native speaker's accent. Many settings present less than ideal listening conditions that have the potential to degrade the communication between non-native speakers and native speakers of a particular language. Because there has been an increase in the number of non-native speakers of English and because English is often recognized as the "international language," the impact of accented speech in less than ideal listening conditions has gained the attention of cross-linguistic researchers.

Although intelligibility of speakers is often influenced by accent, accent alone has been determined to be a poor predictor of intelligibility (Munro & Derwing, 1995a). Heavily accented speech is often rated as highly intelligible by native English speakers. "The amount of information lost [in a message involving a non-native speaker] is presumably related to the type, severity and frequency of divergences from the norms" (Munro & Derwing, 1995b, p. 290). This review has shown that these deviations may include sound substitutions, distortions, and unusual prosodic patterns. In some cases,

these errors are subtle and do not interfere with intelligibility but may require the listener to work harder to understand the message. In other cases, the divergences from native speakers' norms can lead to what a listener perceives as unintelligible speech.

To this point, much of the research has involved comparisons of native listeners' and non-native listeners' perception of a native speaker of English (e.g., Buus et al., 1986; Florentine, 1985; Mayo et al., 1997). Few studies have evaluated the intelligibility of non-native speakers in the presence of noise. For example, the uses of the *SPIN* test have typically involved comparisons between native listeners and non-native listeners responding to recordings of a native American-English speaker in the presence of noise (Florentine, 1985; Mayo et. al., 1997). The *SPIN* has been used in testing the comprehension of English for those who are learning it as a second language; however, until recently (e.g., Schmid & Yeni-Komshian, 1999) it has not been used to test the intelligibility of non-native speakers of English as judged by native listeners or non-native listeners of English. The SPIN test would seem to lend itself to modification for further investigation of the speech intelligibility of non-native speakers of English.

Many studies of speech intelligibility have dealt with native language speakers, hearing impaired speakers, and dysarthric speakers. There is a need to study the intelligibility of non-native speakers in the real-life communication situations they may encounter.

> It is not known whether the effects of noise or filtering (e.g., in telephone or radio transmission, in noisy rooms, or at variable loudness levels) have the same degree of impact on the processing time or comprehensibility of accented speech as on native-produced speech, or whether the effects of such conditions vary as a function of degree of foreign accent. (Munro & Derwing, 1995b, p. 303)

The present study is intended to be a step toward determining factors that cause difficulty for native speakers of English as they attempt to understand non-native speakers in less than ideal environmental conditions. The findings of this research should have implications for people who speak English in situations where it is considered the international language or when English is the common language of two non-native speakers. Noisy backgrounds (multiple sources of noise) and highly intense situations where time is critical can degrade a message that may be intelligible under more ideal conditions. Such settings further complicate the issue of speech intelligibility by presenting additional factors such as noise and lack of visual cues, in addition to the effect of accented speech. The outcome of this research would also contribute to identifying factors that could be adjusted or avoided as preventive safety measures in a high risk, fast paced environment. There are multiple practical applications for this research. One such application might be the facilitation of communication between air traffic controllers and aircrews in international airspace. Another example of a high-risk setting might involve an emergency room situation where the communication occurs in a less than ideal listening environment.

<u>Pilot Study</u>

A pilot study was completed to determine some of the variables and procedures to be considered in investigating the relationship between degree of foreign accent and speech intelligibility of non-native speakers of English with and without contextual cues in the presence of noise. To control for some of the variability between foreign speakers, only native speakers of one language were selected for the pilot study.

Brazilian Portuguese was chosen due to the ease of finding speakers to participate in the project. Speakers who were reliably rated by expert judges to represent mild, mild-moderate, moderate-strong, and strong foreign accents were selected. A native speaker of English was also recorded as a control. The five speakers read sentences from the eight *SPIN* lists. Once the lists were recorded, three lists were randomly chosen. Different levels of multi-talker babble noise were added to each of the three lists representing signal-to-noise ratios of 6 dB, 10 dB, and 15 dB. The target words in half the sentences on each *SPIN* list were classified as highly predictable, whereas the target words in the remaining sentences were of low predictability.

Six adults (3 male and 3 female) who were native speakers of English served as listeners who identified the last word of the sentences produced by speakers with different degrees of accent in all levels of noise. There were a total of three factor levels: two levels of predictability, three signal-to-noise ratios, and five degrees of accent, totaling 30 experimental conditions. The listeners heard four sentences representing each of the 30 conditions for a total of 120 items. The listeners also rated an additional 60- to 90-second connected speech sample from each speaker using rating scales of accent and intelligibility prior to any possible familiarization effect from each listener and again after the experiment to measure potential effects of speaker familiarization.

Preliminary findings from the pilot study indicated that linguistic context (predictability) did contribute to listener accuracy/speaker intelligibility regardless of signal-to-noise ratio or degree of foreign accent. However, the degree of foreign accent also had an impact on speaker intelligibility as measured by listener accuracy. Accuracy

dropped significantly as accent strengthened. There was more variety in accuracy scores among the six listeners as accent increased. Listener accuracy improved as signal-to-noise ratio increased even though differences in performance may not be statistically significant. Differences in listener accuracy between the 6 dB and 15 dB signal-to-noise ratios indicated that accuracy increased as the signal-to-noise ratio increased. Listener accuracy was especially improved in the high predictability context when compared to the low predictability environment at these signal-to-noise ratios. Differences in listener accuracy between the high predictability and low predictability environments decreased by as much as 48-50% as demonstrated in the two strongest accents in the low predictability context.

The strongest influential factors appeared to be a combination of degree of foreign accent and linguistic predictability--whether at the most difficult signal-to-noise ratio of 6 dB or at the most favorable signal-to-noise ratio of 15 dB, listener variability increased and accuracy dropped significantly as a result of an increasingly strong accent in a low predictability context. However, because this observation was not true at the 10 dB level (linguistic context did not make a difference here), it was difficult to draw conclusions from these preliminary findings.

Listeners were less variable and more consistent in their pre- and post-accent ratings of the native and mild accented speakers. Greater variability between raters and between pre- and post- ratings was observed when the accents approached the moderate and strong categories. Listeners rated all speakers as having high intelligibility regardless of the degree of foreign accent. Listeners did not begin to lower their rating of the

speakers until they approached the two speakers with the strongest accent. The lowest rating was 4 out of a possible 5. Listeners remained consistent in their pre- and post-ratings of speech intelligibility, indicating that intelligibility ratings were not significantly influenced by familiarity with the speakers and accents.

Experimental Questions

The primary purpose of this study was to investigate particular contextual (linguistic) and environmental (noise) factors and their relationship to the intelligibility of the speech of non-native speakers of English with varying degrees of accent. A secondary goal was to identify to what degree these factors influence the accuracy of native listeners in understanding non-native speakers of English. Specific questions addressed in this study included the following:

1. Is there a difference in speaker intelligibility based on degree of foreign accent (native, mild, mild-moderate, moderate-strong, strong)?
2. Is there a difference in speaker intelligibility based on signal-to-noise ratio (6 dB, 10 dB, 15 dB)?
3. Is there a difference in speaker intelligibility based on linguistic context (high predictability versus low predictability)?

CHAPTER 2
METHODOLOGY

The purpose of this study was to investigate contextual (linguistic) and environmental (noise) factors and their relationship to the intelligibility of the speech of non-native speakers of English with varying degrees of accent. Data collected from this study were used to answer questions about the degree to which these factors influence native listeners' perceived intelligibility of non-native speakers of English.

Subjects

Speakers

The speech recordings used in this experiment were elicited from four male non-native speakers of English and one native speaker of English. Each speaker represented one category of accent (native, mild, mild-moderate, moderate-strong, or strong). Speaker requirements included:

1. Native speaker of Brazilian Portuguese (with the exception of the one native speaker of English)
2. Males between 18 and 45 years of age
3. No history of a speech disorder
4. No evidence of a current speech disorder as observed by the four expert raters
5. Consistently placed in a particular category of "degree of foreign accent" based on 100% agreement by the expert listeners

The decision to include only male speakers was made to eliminate any possible gender effects. Native speakers of Brazilian Portuguese were selected due to availability and to eliminate any possible differences in responses that might occur if more than one native language were included. All four Brazilian Portuguese speakers selected to participate had passed the Test of English as a Foreign Language (TOEFL), which requires a minimum score of 550 out of total possible score of 660. A preliminary pilot study indicated that the level of English proficiency must be considered. One potential speaker was not included as a participant in the study due to his inability to read the sentences aloud without significant pauses, hesitations, and repeated attempts to pronounce the words. Because poor command of the language may confuse the identification of accentedness, intelligibility, and proficiency by naive listeners, only speakers who were able to read aloud fluently were selected.

Speakers were selected based on a clear differentiation in degree of foreign accent from the other speakers as determined by four expert raters. The expert raters included two experienced teachers of English as a second language and two speech-language pathologists experienced in foreign accent reduction. The pilot study indicated that the speakers could be reliably assigned to four categories: native, mild, moderate, and strong. The expert ratings yielded 100% agreement on those speakers placed in the native, mild, and strong accent categories. However, the expert listeners placed two of the speakers in the moderate category with one speaker rated both at the low end of moderate (3 ratings of 4) and one rating at the high end of mild (1 rating of 3). The other moderate speaker was rated at the high end of moderate (rating of 6) with 100%

agreement. The difference between these two speakers was observed to be sufficient to include both speakers in the study. As a result of these pilot data, the moderate category was, therefore, divided into mild-moderate and moderate-strong.

Listeners

The listeners included 25 males and 25 females from the Gainesville community. Listener requirements included:

1. Native speakers of American-English
2. No significant prior experience listening to speakers of Brazilian Portuguese as NNSs of English
3. Between the ages of 18 and 40 years
4. Hearing abilities appropriate for the task

The decision to include only native speakers of American-English was made to eliminate any possible differences in listener responses that might occur if more than one native language was included. A questionnaire was presented to each listener to determine if the listener had been exposed to Brazilian Portuguese speakers and to other foreign accents (Appendix A). Gass and Varonis (1984) reported that prior experience with non-native speech in general and even familiarity with a particular nonnative accent facilitates the comprehension of the speech of another non-native speaker of that language background. For this reason, listeners with prior exposure to Brazilian Portuguese speakers were not included in this study. In the pilot study, one listener was fluent in Russian as a second language and was familiar with Russian speakers of English. However, his accuracy on the listening task was similar to the other listeners who reported no significant exposure to speakers with a foreign accent.

Listeners between the ages of 18 and 40 years were selected due to availability and also to rule out the possible effects of aging on hearing acuity and speech discrimination. Hearing ability appropriate for the task was determined by a 92% or better performance on the Griffiths Intelligibility Test of Speech Discrimination (Griffiths, 1967). The Griffiths Intelligibility Test was administered to groups of potential listeners in a sound-treated room at a comfortable listening level agreed upon by the listening group. The test is based on the CID W-22 word lists especially developed for the purpose of assessing speech discrimination. The listeners were required to circle the word presented via audio recording from a list of five written words that are phonemically different by just one consonant. The vowel remains constant among the five choices. Each list of five stimuli varied either in initial consonant or final consonant but not both (Appendix B).

Experimental Stimuli

Two randomized digital audio-tape (DAT) recordings of 210 sentences selected from the *SPIN* lists and representing each speaker in each of the noise and linguistic conditions were developed as the experimental stimuli for the assessment of speech intelligiblity. The audio-tapes also included an additional 60- to 90-second speech sample of each speaker for pre- and post-evaluation ratings of accent and intelligiblity.

The *SPIN* sentences were selected as the experimental stimuli for several reasons. They were developed specifically to study the intelligibility of speech in noise (Kalikow et al., 1977) and have been used in several studies with normal hearing, hearing-impaired, and non-native listeners (e.g., Dirks et al., 1981; Elliott, 1979; Florentine, 1985; and

Owens, 1981). Contextual speech intelligibility tests have shown more external validity as measures of real world speech intelligibility than isolated word intelligibility tests (Schiavetti et al.,1984). The *SPIN* test was designed to assess a listener's speech recognition while using linguistic-situational information in speech. The listener hears a recording of a list of sentences presented with a background of multi-talker babble and repeats or writes the last monosyllable (target word) of each sentence. Each of the 50 sentence lists contains 25 high predictability target words and 25 low predictability target words.

For this study each speaker was recorded reading all eight of the *SPIN* lists (Appendix C). Later, three of the lists were chosen randomly and mixed with the multi-talker babble noise with an assigned signal-to noise ratio (SNR) of 6 dB, 10 dB, or 15 dB. After mixing, seven high predictability sentences and seven low predictability sentences from each of the three lists were randomly selected via a computer-generated randomizing program. These 210 sentences (14 sentences x 3 lists x 5 speakers) were randomized again via the randomizing program. The randomly selected sentences were generated by dubbing from the mixed recordings in the order generated by the program to create the two audio-tapes for the final listening task. The two randomized versions were generated to reduce the potential order effect associated with increased listener familiarity (Appendix D).

Recording/Instrumentation

Individual recording sessions were conducted in a sound-treated audiometric booth with a Sony-Digital Audio Tape Deck, (Model DTC 690) using a preamplifier,

(Model DBX 707) and an Audio-Technica microphone, (Model ATM73a) attached to a light plastic headset and positioned at a constant one inch from the right corner of the speaker's lips. The speakers were provided with type-written sentences from the *SPIN* lists 2.1 through 2.8. Speakers were allowed to review the sentences prior to recording and to ask any questions for clarification. Each list was recorded in a single recording session.

At a later time the recordings were edited and mixed with the selected level of noise. Each list was played through one channel from a Sony DAT, (Model 59ES), while the multi-talker babble was played through the other channel from a Sony Cassette, (Model RX 606ES). The outputs of the two channels were routed to a speech audiometer amplifier/attenuator system (Grason-Stadler, Model GSI 16) where the signal and noise tracks were mixed. As a result, the speech signal and noise were mixed and recorded into both channels simultaneously. A 1000-Hz calibration tone was recorded on the signal and noise tracks to monitor and adjust the speech and noise signals prior to recording to provide equivalent signal levels across the recorded speech samples (Morgan et al., 1981). Once the master tapes were made with the appropriate SNR corresponding to the selected list (i.e., SPIN List 2.1 had a SNR of 6 dB; SPIN List 2.3 had a SNR of 10 dB; and SPIN List 2.6 had a SNR of 15 dB), seven high predictability sentences and seven low predictability sentences from each list were randomly selected.

Procedure

This researcher conducted each test session to provide consistency of instruction and procedure. The listening task was presented to groups of two to eight listeners at a time in a quiet sound-treated room. On three occassions, the task was administered to one listener due to cancellations and scheduling conflicts of other volunteers. Playback was via a Sony Digital Audio Tape-Corder, (Model TCD-D8) through Altec Lansing Mulitmedia Computer Speaker System, (Model ACS90). Listeners were provided with definitions and instructions both verbally and in a written format (Appendix E). They were encouraged to ask questions to clarify any confusion in terminology or procedure. The entire session took approximately 60 minutes. A short break was offered to the listeners approximately half-way through the session.

Most Comfortable Loudness Level (MCL)

The MCL was established in the sound-treated room and with the stimuli to be used in the experiment at a predetermined playback level of 60-65 dB. Once the MCL was determined by consensus, the loudness level remained constant for each set of listeners. Listeners were instructed to indicate whether the playback was too soft or too loud and it would be adjusted accordingly. In all 16 sessions, listeners were satisfied with the level of playback at the predetermined MCL.

Degree of Accent Rating

Once the instructions were reviewed and the MCL established, listeners were asked to listen to a 60- to 90-second spontaneous speech sample from each speaker and to rate degree of foreign accentedness on a 10-point Likert scale with 0 representing no

detectable foreign accent, 1-3 representing a mild accent, 4-6 a moderate accent, and 7-9 a strong accent (Appendix F).

Intelligibility Rating

Listeners were also instructed to rate the level of overall intelligibility of each speaker on a 5-point Likert scale with 1 representing no comprehension at all, 2 representing considerable difficulty understanding the speaker with a listener only able to pick out single words or a few phrases, 3 representing comprehension of approximately 50% of the speech sample, 4 representing comprehension of most of the speech sample with the exception of a few words or phrases, and 5 representing comprehension of 98-100% of the message. This intelligibility rating was used to establish a baseline rating prior to any effects of familiarization (Appendix F).

Presentation of the *SPIN* Test Stimuli

Listeners were familiarized with the *SPIN* task by listening to five trial items presented at a 20 dB SNR. One item from each speaker was taken from an unused *SPIN* list. The listeners were provided with a numbered form (Appendix G) and instructed to write down the last word of each sentence corresponding to the appropriate number. Listeners were encouraged to guess if they were unsure of any words and instructed to fill in all blanks on the listener response form.

Post-Listening Ratings

At the end of the listening session, listeners were asked to provide a final rating of accentedness and intelligibility for all speakers using the same recordings previously

presented to and rated by each listener. This post-task rating was compared with initial ratings to determine any effects of familiarization.

Scoring

Each 210-sentence form was scored by the researcher based on word recognition scores. To be scored as correct, each response clearly represented the exact word, allowing spelling errors but not morphological errors. Questionable responses (i.e., *den-* -> *dean*) were submitted to a second scorer. Results were analyzed based on the 30 listening conditions: the five degrees of accent, three levels of noise, and the two levels of predictability.

Statistical Analysis

The data were analyzed using a three-way factorial treatment design in a randomized block design. Listener effect was treated as random effect while the degree of foreign accent (native, mild, mild-moderate, moderate-strong, strong), level of noise (6 dB, 10 dB, 15 dB SNRs), and linguistic predictability (low, high predictability) were treated as fixed effects. The treatments were the combinations of the levels of the three factors: accent, noise, and predictability.

The advantage of a factorial design lies in its ability to test whether the effect of one factor depends on the level of the other factor (first order interaction). The factorial design also tests whether two-way interactions (second order interaction) depend on the levels of the third factor.

The data were statistically analyzed using an Analyis of Variance (ANOVA) procedure. The analysis was performed using SAS General Linear Models (GLM) for

Mixed Models. The ANOVA procedure allows for analysis of mulitple factors with any number of levels and permits detection of interactions among the various factors (Maxwell & Satake, 1997). This study involved three factors that included more than one level to be analyzed. That is, there were three levels of noise, two levels of predictability, and five levels of accent to be analyzed, totaling 30 combinations of the three factors of interest (3 x 2 x 5). Seven response variable measurements were obtained under each factor level combination for a total of 210 (3 x 2 x 5 x 7) response variable measurements for each listener (50). The 30 conditions multiplied by the 50 listeners in this study resulted in a total of 1500 (30 x 50) data points that were analyzed or 10,500 (7 x 30 x 50) measurements of listener accuracy.

CHAPTER 3
RESULTS

This study investigated particular contextual (linguistic) and environmental (noise) factors and their relationship to the intelligibility of the speech of non-native speakers of English with varying degrees of accent. A secondary goal identified the degree to which these factors influenced the accuracy of native listeners in understanding non-native speakers of English. The study was designed to examine the following independent variables: linguistic context, noise, and degree of foreign accent. The dependent variable was the number of correct responses provided by each listener in each of the above conditions.

The data were analyzed using a 5 x 3 x 2 (Degree of foreign accent x Level of Noise x Predictability) factorial treatment design in a randomized block experimental design (listeners being the blocks). Listener effect was treated as random effect while the degree of foreign accent (native, mild, mild-moderate, moderate-strong, strong), level of noise (6 dB, 10 dB, 15 dB signal to noise ratio) and linguistic predictability (low or high predictability) were treated as fixed effects. The treatments were the combinations of the levels of the three factors: accent, noise, and predictability. The statistical analysis was performed using SAS General Linear Models (GLM) Analysis of Variance for Mixed Models. Although all main effects and two-way interactions reached

statistical significance, the ANOVA detected a statistically significant three-way interaction between degree of foreign accent, level of noise, and linguistic predictability, $F(8,1421) = 21.06, p = .0001$. Because the three factors interact, evaluating them separately would hide the interaction. The combination of the three factors is needed for true evaluation of any differences. Table 3-1 provides the ANOVA table with the relevant statistics and probability values for each factor. Statistical significance was established at $p < .01$ for all analyses reported throughout this study.

Table 3-1. Analysis Of Variance Examining the Effects of Listener, Accent, Noise, Predictability, and Interactions.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Listener	49	214.5340	4.3782	6.07	0.0001*
Accent	4	1191.3973	297.8493	413.21	0.0001*
Noise	2	710.2240	355.1120	492.65	0.0001*
Accent x Noise	8	124.1027	15.5128	21.52	0.0001*
Predictability	1	1861.4940	1861.4940	2582.47	0.0001*
Accent x Predict	4	143.5893	35.8973	49.80	0.0001*
Noise x Predict	2	89.7760	44.8880	62.27	0.0001*
Accent x Noise x Predic	8	121.4307	15.1788	21.06	0.0001*
Error	1421	1024.2860	0.7208		

*statistically significant at $p < .01$

The ANOVA (Table 3-1) showed a statistically significant listener effect, $F(49, 1421) = 6.07, p = .0001$, indicating that blocking increased information in the experiment. In other words, there was a significant difference between listeners and variability was reduced by treating the listeners as blocks, providing a more precise estimate of error.

The results of this study are presented relative to the experimental questions posed. First, is there a difference in speaker intelligibility based on degree of foreign accent (native, mild, mild-moderate, moderate-strong, strong)? Second, is there a difference in speaker intelligibility based on signal-to-noise ratio (6 dB, 10 dB, 15 dB)? And third, is there a difference in speaker intelligibility based on linguistic context (high predictability versus low predictability)? For each question the results of the statistical test will be presented first, followed by a description of the means.

The Effect of Degree of Accent

Each listener had 42 opportunities to hear each accent. Each accent was heard in seven low predictability sentences and seven high predictability sentences at each of the three noise levels (14 x 3 = 42).

The significant three-way interaction, $F(8,1421) = 21.06, p =.0001$, indicates that the effect of level of noise and linguistic predictability on intelligibility depends on the degree of foreign accent. As a result of the three-way interaction, a contrast analysis of the interaction between noise level and linguistic predictability was investigated separately for each degree of foreign accent. Table 3-2 contains the relevant statistics and probability values of the tested contrasts.

A statistically significant first order interaction was detected between noise level and predictability for the native, mild, mild-moderate, and moderate-strong accent levels ($p < .0001$) but not for the strong accent level, $F(2, 1421) = 4.46, p =.0117$. These values indicate that the effect of predictability on intelligibility depends on the noise level, and that the noise level effect depends on the predictability level of each degree of

foreign accent. The interaction between noise levels and predictability levels was the weakest for the strongest accent. In general, listeners were less accurate as noise level increased, as degree of foreign accent increased, and when predictability was low. Therefore, degree of foreign accent did make a difference in speaker intelligibility, and as degree of foreign accent became stronger the intelligibility became poorer.

Table 3-2. Table of Contrasts Investigating One-Way Interactions (Inter) Between Noise Level and Predictability on Degree of foreign accent Levels.

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Inter Native	2	20.9267	10.4633	14.52	0.0001*
Inter Mild	2	48.7467	24.3733	33.81	0.0001*
Inter Mild-Moderate	2	92.5800	46.2900	64.22	0.0001*
Inter Moderate-Strong	2	42.5267	21.2633	29.50	0.0001*
Inter Strong	2	6.4267	3.2133	4.46	0.0117
Error	1421	1024.2860	0.7208		

*statistically significant at $p < .01$

The means for all correct listener responses were examined and are shown in Table 3-3 and illustrated graphically in Figures 3-1 and 3-2. As seen in Figure 3-1, the most difficult listening condition (6 dB SNR, LP, indicated by diamond markers) was a consistent problem for all accents, even when listening to the native speaker. Greater noise levels (square and triangular markers) presented progressively more difficulty for listeners as degree of foreign accent became stronger in the low predictability condition. Although there was some unevenness to this pattern in the high predictability condition (Figure 3-2), in general, the trend remained similar.

Listeners demonstrated the smallest differences in the means between the low predictability and high predictability conditions with the native accent (Table 3-3). The general trend is that differences in listener accuracy became progressively greater as degree of foreign accent increased with the exception of the strong accent presented in a SNR of 15 dB and the moderate-strong and strong accents at 6 dB. These trends can also be seen in Figures 3-3, 3-4, and 3-5.

Table 3-3. Means (M), Standard Deviations (SD), and Mean Differences Between Low (LP) and High (HP) Predictability Targets of All Listeners' Responses in Each Condition

	Degree of Accent									
	Native		Mild		Mild-Mod		Mod-Strong		Strong	
	M	SD	M	SD	M	SD	M	SD	M	SD
6 dB										
LP	4.44	0.99	3.10	1.31	2.26	1.10	2.10	1.43	1.30	1.07
HP	6.42	0.86	5.96	0.88	6.46	0.68	4.46	1.09	4.48	1.40
Mean Differences	1.98		2.86		4.20		2.36		3.18	
10 dB										
LP	5.92	0.44	5.44	0.86	4.52	1.07	2.42	1.16	3.60	1.20
HP	6.76	0.43	6.94	0.24	6.68	0.51	4.44	1.28	6.18	1.04
Mean Differences	0.84		1.50		2.16		2.02		2.58	
15 dB										
LP	6.08	0.85	5.98	1.00	5.20	0.99	2.44	0.58	4.06	0.87
HP	6.96	0.20	6.92	0.27	6.82	0.39	6.20	0.73	6.60	0.57
Mean Differences	0.88		0.94		1.62		3.76		2.54	

Note. Total possible in each condition = 7.0

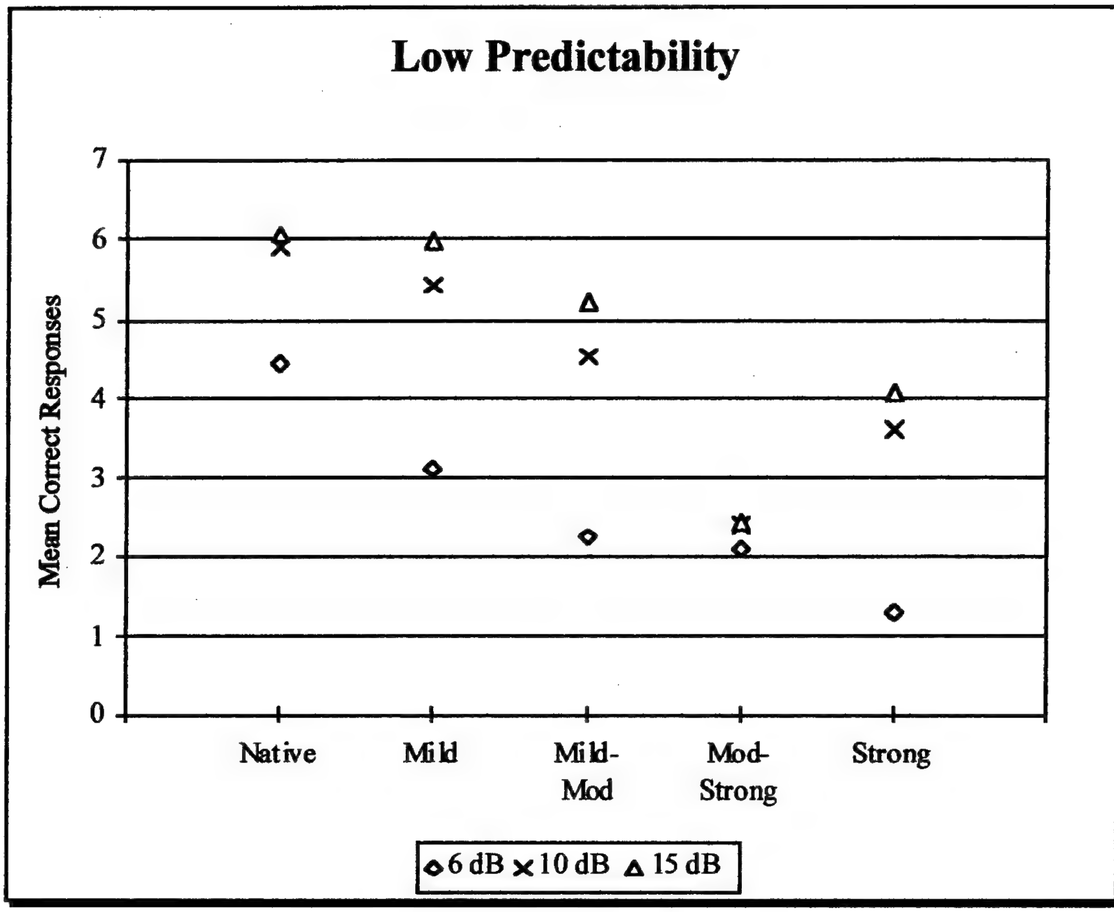

Figure 3-1. Mean Correct Responses for All Listeners at the Low Predictability Level.

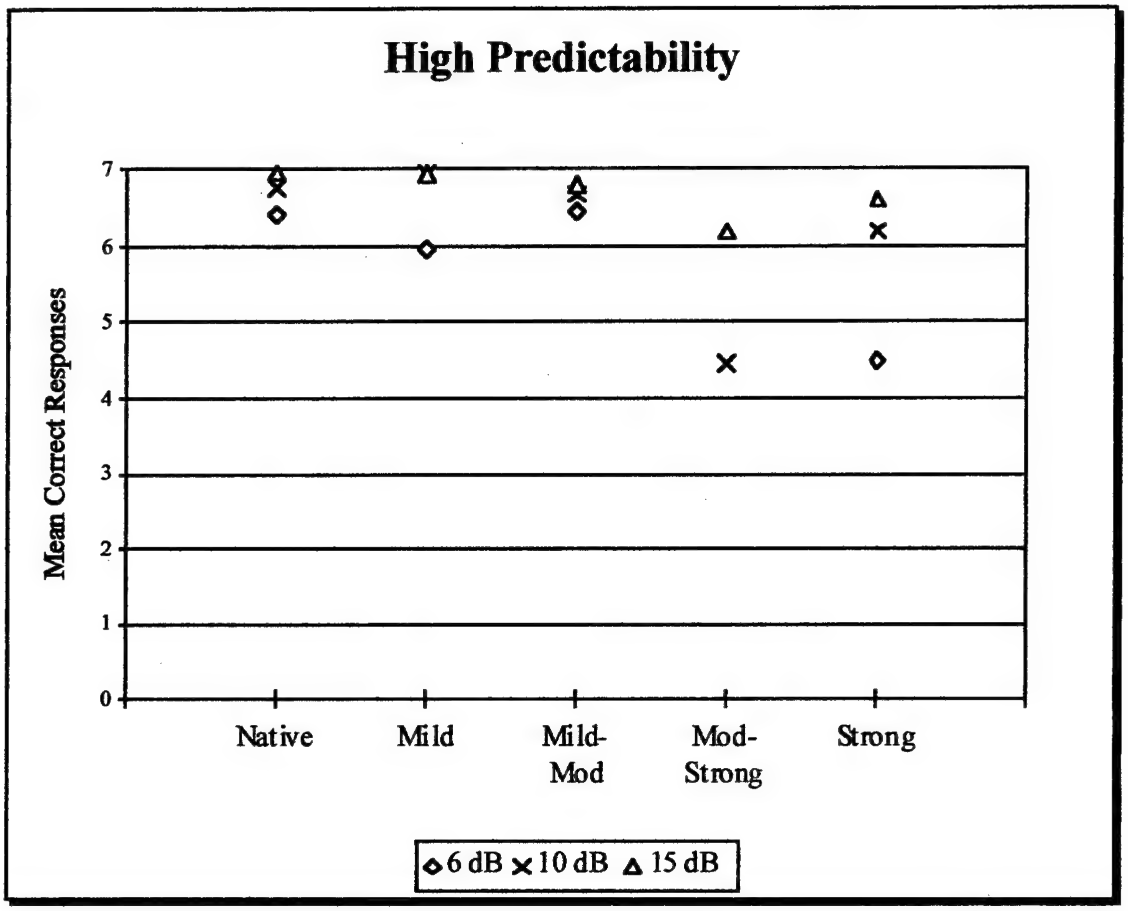

Figure 3-2. Mean Correct Responses for All Listeners at the High Predictability Level.

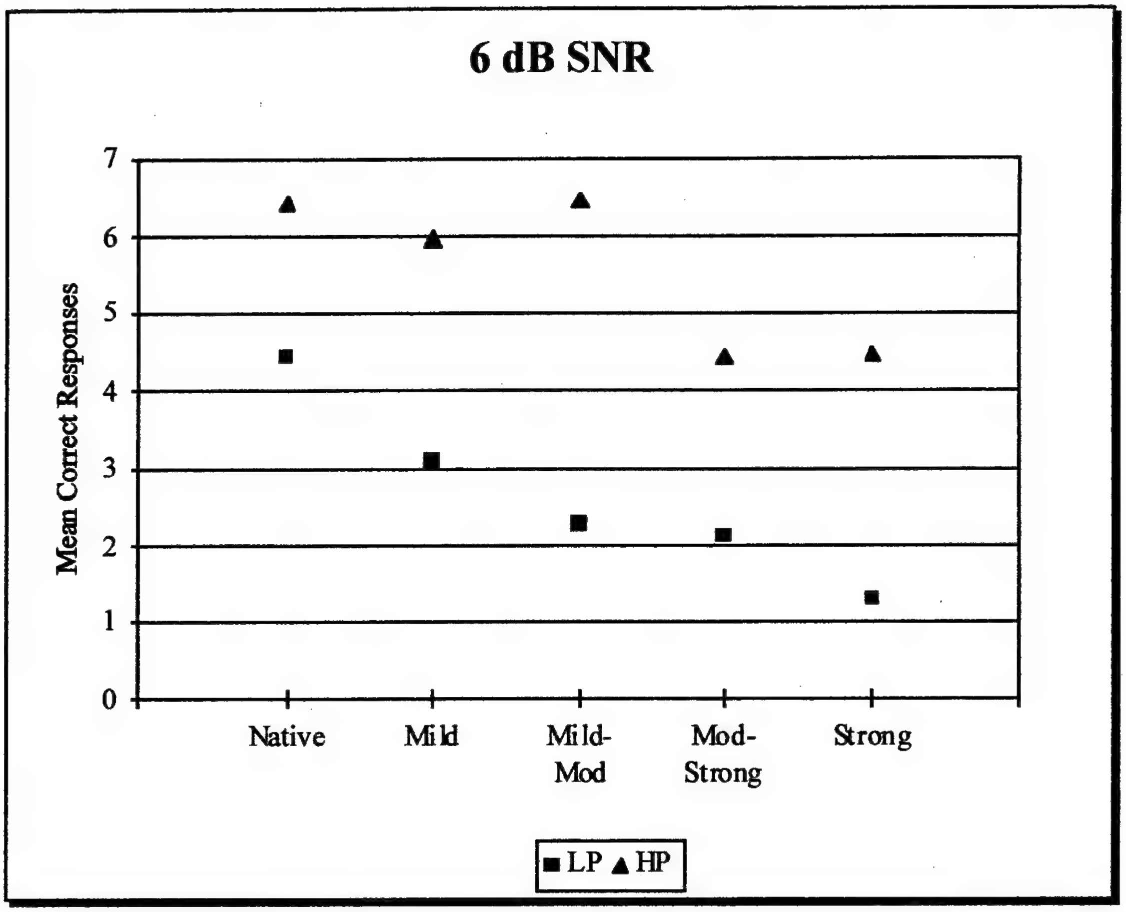

Figure 3-3. Mean Correct Responses for All Listeners at 6 dB SNR at High (HP) and Low (LP) Predictability Levels.

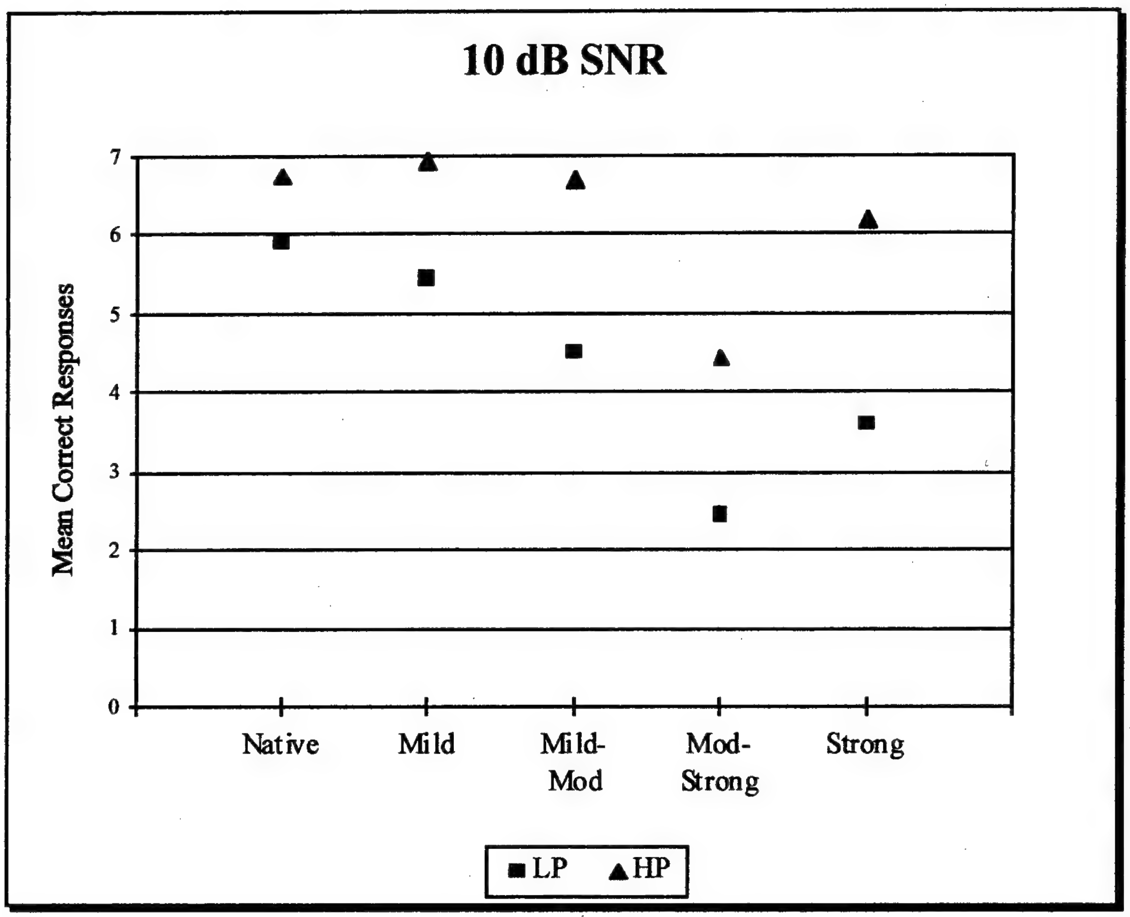

Figure 3-4. Mean Correct Responses for All Listeners at 10 dB SNR at High (HP) and Low (LP) Predictability Levels.

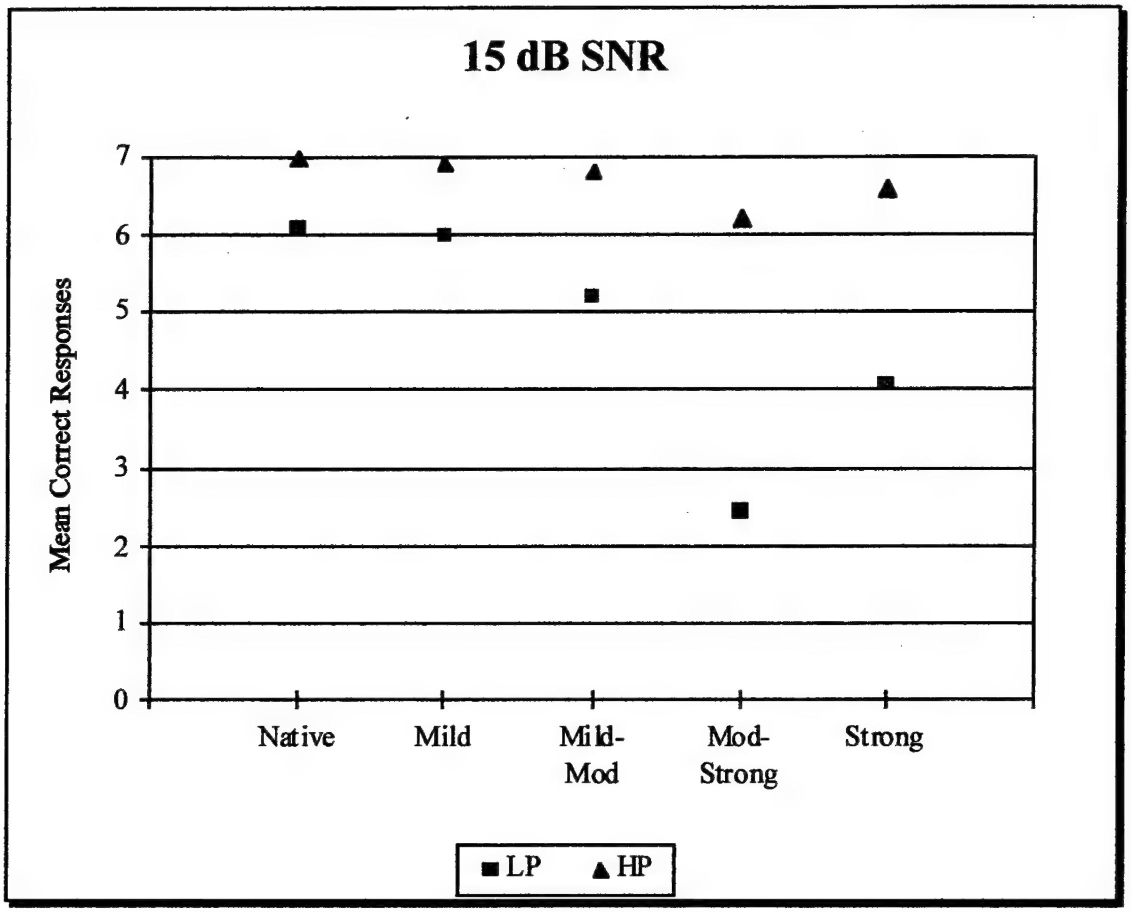

Figure 3-5. Mean Correct Responses for All Listeners at 15 dB SNR at High (HP) and Low (LP) Predictability Levels.

The Effect of Signal to Noise Ratio

Each listener had 70 opportunities to hear sentences at each noise level. Each noise level was heard in seven low predictability sentences and seven high predictability sentences at each of the five levels of accent (14 x 5 = 70).

The effect of noise level was investigated for each degree of foreign accent and predictability level combination. As seen in Table 3-4, a statistically significant ($p < .01$) noise level factor effect was detected for native, mild, and strong degrees of accent on both predictability levels. A statistically significant noise level effect was detected for the mild-moderate accent at the low predictability level ($p < .01$) but not at the high predictability level ($p = .1022$). On the other hand, the noise level effect was not significant at the low predictability level ($p = .0804$) with the moderate-strong accent but was significant at the high predictability level.

The mean differences in listener responses between the low and high predictability conditions (Table 3-3) at the extreme noise levels of 6 dB and 15 dB became progressively smaller as SNR increased except with the moderate-strong speaker. This trend was not as clear at the 10 dB level. Listeners did not demonstrate greater accuracy for the moderate-strong speaker regardless of noise level in the low predictability condition (Figure 3-1). In fact, the mean range of accuracy was between 2.12 and 2.44 out of 7.0 at any noise level. Although listener accuracy was greater in the high predictability than the low predictability conditions at 6 and 10 dB SNRs, when listeners heard the moderate-strong speaker at 15 dB in the high predictability context,

mean accuracy was even greater (6.20). This speaker was highly intelligible when listening conditions were optimal (Figure 3-2).

Table 3-4. Table of Contrasts Investigating Noise Effect for Each Degree of Foreign Accent and Predictability Level.

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Noise Native LP	2	81.7600	40.8800	56.71	0.0001*
Noise Native HP	2	7.4533	3.7267	5.17	0.0058*
Noise Mild LP	2	234.3600	117. 1800	162.56	0.0001*
Noise Mild HP	2	31.3733	15.6867	21.76	0.0001*
Noise Mild-Mod LP	2	236.8933	118.4467	164.32	0.0001*
Noise Mild-Mod HP	2	3.2933	1.6467	2.28	0.1022
Noise Mod-Str LP	2	3.6400	1.8200	2.52	0.0804
Noise Mod-Str HP	2	102.0933	51.0467	70.82	0.0001*
Noise Strong LP	2	218.6533	109.3267	151.67	0.0001*
Noise Strong HP	2	126.0133	63.0067	87.41	0.0001*
Error	1421	1024.2860			

*statistically significant at $p < .01$

The Effect of Linguistic Predictability

Each listener listened to 105 low predictability sentences and 105 high predictability sentences for a total of 210 items. Each speaker (5 levels of accent) was heard in seven low predictability sentences and seven high predictability sentences at each of the three noise levels (14 x 5 x 3 = 210).

The effect of linguistic predictability was investigated separately for each degree of foreign accent and noise level combination. A statistically significant predictability effect was detected on each noise level for every degree of foreign accent ($p = .0001$). Table 3-5 contains relevant statistics and probability values of the tested contrasts.

Table 3-5. Table of Contrasts Investigating Predictability (Pred) Effect for Each Degree of foreign accent and Noise Level.

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Pred in Native 6dB	1	98.0100	98.0100	135.97	0.0001*
Pred in Native 10dB	1	17.6400	17.6400	24.47	0.0001*
Pred in Native 15dB	1	19.3600	19.3600	26.86	0.0001*
Pred in Mild 6dB	1	204.4900	204.4900	283.69	0.0001*
Pred in Mild 10dB	1	56.2500	56.2500	78.04	0.0001*
Pred in Mild 15dB	1	22.0900	22.0900	30.65	0.0001*
Pred in Mild-Mod 6dB	1	441.0000	441.0000	611.80	0.0001*
Pred in Mild-Mod 10dB	1	116.6400	116.6400	161.82	0.0001*
Pred in Mild-Mod 15dB	1	65.6100	65.6100	91.02	0.0001*
Pred in Mod-Strong 6dB	1	139.2400	139.2400	193.17	0.0001*
Pred in Mod-Strong 10dB	1	102.0100	102.0100	141.52	0.0001*
Pred in Mod-Strong 15dB	1	353.4400	353.4400	490.33	0.0001*
Pred in Strong 6dB	1	252.8100	252.8100	350.73	0.0001*
Pred in Strong 10dB	1	166.4100	166.4100	230.86	0.0001*
Pred in Strong 15dB	1	161.2900	161.2900	223.76	0.0001*
Error	1421	1024.2860	0.7208		

*statistically significant at $p < .01$

As reported previously, a statistically significant noise level factor effect was detected for the native, mild, and moderate-strong accents at *both* predictability levels.

When examining the effect of predictability (Table 3-4), the mild-moderate accent did show a statistically significant noise level effect $F(2, 1421) = 164.32, p < .0001$at a low predictability level but not in the high predictability level, $F(2, 1421) = 2.28, p = .1022$. In contrast, the moderate-strong accent demonstrated a significant difference $F(2, 1421) = 70.82, p < .0001$ at the high predictability level but not in the low predictability level across all noise levels, $F(2, 1421) = 2.52$, ($p = .0804$).

Pre- and Post-Task Ratings

Listener Ratings of Degree of Accent

Prior to the SPIN listening task as well as following the listening task, listeners were asked to rate degree of foreign accent based on a 60- to 90-second connected speech sample in which each speaker talked about a subject of his choice. The four categories for degree of foreign accent were presented in a 10-point Likert scale (Appendix E) with 0 representing no detectable accent, 1-3 representing a mild accent, 4-6 representing moderate, and 7-9 representing a strong accent.

It can be seen in Table 3-6 that a statistically significant difference was detected in accent ratings prior to and after the listening session for mild and strong degrees of accent ($p = .0001$), indicating that familiarization with the accents and task had an effect on the listener ratings for these two degrees of accent. No statistically significant differences were detected for the mild-moderate ($p = .5486$) or for the moderate-strong accent ($p = .0292$). Table 3-6 contains the differences between the means (Post - Pre), the standard deviation of the differences, and the probability values for each accent level.

Table 3-6. Ratings for Accent: Means, Standard Deviations, Differences Between the Means (Post-Pre), Standard Deviation of the Differences, and p-values for Each Accent Level for All Listeners Pre- and Post- Listening Task (Scale: 0-9).

	Degree of Accent							
	Mild		Mild-Mod		Mod-Strong		Strong	
	M	SD	M	SD	M	SD	M	SD
Pre	2.24	1.33	4.08	1.66	4.90	1.72	5.44	1.55
Post	3.16	1.60	4.20	1.75	5.26	1.56	6.30	1.39
Mean of the differences (post-pre)	0.92	1.19	0.12	1.42	0.36	1.17	0.86	1.14
p-value	0.0001*		0.5486		0.0292		0.0001*	

Note. Degree of foreign accent: 0=No detectable accent; 1-3=Mild; 4-6=Mod; 7-9=Strong.
*statistically significant at $p < .01$

Listeners were 98% accurate in their identification of the native speaker in the pre- and 100% accurate in the post-task ratings (Appendix H). The means of listener ratings demonstrated that listeners placed the speakers in the same sequential order from mild to strong as the expert raters (Table 3-6). The means reflect that as a group, the listeners assigned a slightly stronger accent in the post-task rating than in the pre-task rating. This pattern was consistently seen across all speakers, although this difference was signficant only for the mild and strong accents.

Listener Ratings of Intelligibility Level

Listeners were asked to rate the intelligibility of each speaker from the same connected speech sample pre- and post-task. Definitions were provided to clarify the

differences between accent and intelligibility to guide the listener in rating the appropriate feature. The rating scale for intelligibility was based on a 5-point Likert scale with 1 representing no comprehension of the speaker at all and 5 representing 98-100% comprehension of the entire message (Appendix E). No statistically significant difference was detected in intelligibility ratings pre- or post- listening session for any accent degree ($p > 0.197$), showing that familiarization with the accents and task had no effect on the listener ratings of intelligibility. Table 3-7 contains the differences between the means (Post and Pre), the standard deviation and probability values for each accent level. The ratings of all 50 listeners were averaged to compare pre- and post-ratings. The data in Table 3-7 reveal that the listeners rated all speakers as having high intelligibility regardless of the degree of foreign accent. Because intelligibility was rated very high in the pre-task rating for all speakers, an effect of familiarization could not be detected. Post-task intelligibility ratings were extremely close to the pre-task ratings. Although intelligibility ratings were high for all speakers, the mean ratings for each speaker again followed the degree of foreign accent assigned to the speakers by the expert raters.

Table 3-7. Ratings for Intelligibility: Means, Standard Deviations, Differences Between the Means (Post-Pre), Standard Deviation of the Differences, and p-values for Intelligibility for All Listeners Pre- and Post- Listening Task (Scale: 1-5).

	Degree of Foreign Accent							
	Mild		Mild-Mod		Mod-Strong		Strong	
	M	SD	M	SD	M	SD	M	SD
Pre	4.98	0.14	4.94	0.24	4.72	0.50	4.34	0.66
Post	4.96	0.20	4.92	0.27	4.72	0.45	4.42	0.57
Mean of the differences (post-pre)	0.02	0.25	0.02	0.14	0.00	0.29	0.08	0.44
p-value	0.5686		0.3124		1.0000		0.1970	

Note. Intelligibility Rating: 1=I did not understand the speaker at all. 2=I had a lot of difficulty understanding the speaker--I could only pick out single words or a few phrases. 3=I was able to understand about 50% of the speech sample. 4=I understood most of the sp sample with the exception of a few words or phrases. 5=I understood 98-100% of the entire message.

CHAPTER 4
DISCUSSION

The results of this study suggest that all three factors--accent, noise, and predictability--had a combined effect on the perceived intelligibility of non-native speakers when judged by native speakers of American English. In fact, even the intelligibility of the native speaker was compromised when the signal-to-noise ratio was low and when the linguistic predictability was also low. However, when the native listeners were placed in the same condition but challenged further by the addition of a foreign accent, intelligibility was even more compromised. This effect is greater as the degree of accent became progressively stronger.

The findings of this study will be discussed in order of the experimental questions posed. First, is there a difference in speaker intelligibility based on degree of accent (native, mild, mild-moderate, moderate-strong, strong)? Second, is there a difference in speaker intelligibility based on signal-to-noise ratio (6 dB, 10 dB, 15 dB)? And third, is there a difference in speaker intelligibility based on linguistic context (high predictability versus low predictability)? For each question, the discussion that follows will include related studies previously reported in the review of the literature, interpretations drawn from the statistical analysis, and possible explanations for the patterns observed.

For the purposes of this discussion, the term "practical significance level" will be used to designate conditions in which the mean listener accuracy (responses correct) was less than 6 out of a possible 7. This means that out of seven possible target words in each condition, listeners were allowed to miss an average of one word, or about 15% before it was interpreted as having practical significance. Misunderstanding 15% of the target words in a real-life situation where listening conditions are less than ideal and precise communication is extremely critical in a fast-paced, high-risk environment will be considered to pose a potentially serious risk (Table 4-1).

Table 4-1. Mean Percent Correct--Rounded to the Closest Whole Percentage Point--for All Listeners' Responses in Each Condition.

	Degree of Foreign Accent				
	Native %	Mild %	Mild-Mod %	Mod-Strong %	Strong %
6 dB					
LP	63	44	32	30	19
HP	92*	85*	92*	64	64
Mean Differences	29	41	60	34	45
10 dB					
LP	85*	78	65	35	51
HP	97*	99*	95*	63	88*
Mean Differences	12	21	30	28	37
15 dB					
LP	87*	85*	74	35	58
HP	99*	99*	97*	89*	94*
Mean Differences	12	14	23	54	36

* greater than 85%

The Effect of Degree of Accent

The effect of degree of accent on speaker intelligibility was examined in two ways. First, speaker intelligibility was measured by asking listeners to identify the last word of a sentence selected from the *SPIN* Test. The number of correct listener responses was then used as a measure of speaker intelligibility. The second measure of intelligibility was made by asking listeners to rate both intelligibility and degree of accent on Likert scales. The relationship between perceived accent and perceived intelligibility was then compared pre- and post-task. The main purpose of the pre- and post-task comparison was to determine if there was a familiarization effect that would influence listeners' judgments once they had been exposed to these speakers. Both methods of intelligibility measurement have been frequently used in the study of speech intelligibility of non-native speakers of English.

The findings of this study are in agreement with those of Schmid and Yeni-Komshian (1999) who used the *SPIN* sentences produced by both native and non-native speakers of English to investigate the processing effort associated with recognizing and comprehending accented speech in comparison to native-sounding speech. They concluded that even though non-native speakers are judged as intelligible, listeners may have to expend more effort to recognize and comprehend accented speech in comparison to native-sounding speech. They also found that degree of accent had an effect on listener accuracy. That is, listeners more accurately identified mispronunciations embedded in sentences when listening to mildly-to-moderately accented speakers than when listening to heavily accented speakers.

The listeners in the current study were more accurate when listening to the mild accent and their accuracy decreased as accent strengthened. This was seen especially in the low predictability context (Figure 3-1). In the high predictability context, listeners demonstrated high levels of accuracy with all speakers in the best noise condition (15 dB SNR); however, listener accuracy decreased substantially with the two strongest accents in the worst noise condition (6 dB SNR) (Figure 3-2).

The findings of this study support those of Munro and Derwing (1995a) who found that listeners tended to rate accent more harshly than they rated comprehensibility. In their study, accent scores were a poor reflection of what listeners actually comprehended. In many cases, non-native speakers of English were rated as moderately or heavily accented, but listeners were able to transcribe the messages perfectly. Munro and Derwing concluded that the presence of a strong foreign accent does not necessarily result in reduced intelligibility. The findings of this study are consistent with those of Munro and Derwing. That is, listeners rated all speakers as having high intelligibility regardless of degree of accent (Table 3-7). Degree of accent, on the other hand (Table 3-6), was rated more harshly than intelligibility. Even though many individual listeners used the full 10-point scale when rating accent, they still rated the strongest speaker on the top end of the 5-point Likert scale for intelligibility. This imbalance between ratings of accent and intelligibility may be explained in part by subject selection. The speakers selected for this study had passed the *TOEFL* exam for the purposes of entrance to graduate school. Although the *TOEFL* does not provide a measurement of pronunciation, it does assure a minimal level of English proficiency.

This speaker criterion was selected in order to prevent complications in measuring accentedness and intelligibility with the issue of a poor command of the language. This selection of speakers, then, represents the high end of the spectrum of non-native speakers. Situations that involve non-native speakers of English may very well include speakers who are less proficient with the language and may present distractions in the message such as significant pauses, hesitations, and repeated attempts to pronounce the words. These additional distractions may further reduce speaker intelligibility. The findings of this study, then, focus on those speakers who represent a range of accents but who have a relatively good command of the language. It is likely that listeners would have greater difficulty with speakers who are less proficient in English.

It is interesting that the listeners placed the speakers in the same sequential order from mild to strong accent as the expert raters, showing agreement among expert raters and naive listeners in their ability to categorize speakers according to accent. Surprisingly, listeners rated accent more harshly in the post-task rating than they did in the pre-task rating; however, the sequential order of degree of accent remained consistent. The harsher ratings of accent in the post-task ratings suggest that listeners were less tolerant of accent after becoming familiar with all speakers. The literature would predict ratings would improve as a result of familiarity. One explanation of the opposite findings in this study could be that when listeners did the pre-task ratings, they had not yet heard the full range of speakers and were more conservative in their ratings. However, by the time they completed the post-task ratings, they were familiar with the full range of speakers and more comfortable with the full use of the scale. It is

also possible that listeners assigned stronger ratings of accent due to fatigue or anxiousness to finish the session. It could be that the listeners were less cautious and less interested as a result of the placement of this task at the end of the experiment.

In summary, the findings of this study do suggest that there is a difference in intelligibility of non-native speakers of English based on degree of accent. Although the effects of noise and predictability have an impact on intelligibility, the effects of these conditions seem to vary as a function of degree of accent.

The Effect of Signal to Noise Ratio

Early research summarized by Denes and Pinson (1993) involving the effect of noise on speech intelligibility generally referred to a white noise background and word articulation scores based on the percentage of words correctly identified in a list of phonetically balanced words. Denes and Pinson reported that the impact of white noise at a 20 dB signal-to-noise ratio had no effect on speech intelligibility but a 0 dB signal-to noise-ratio would yield a word articulation score of 50%. Nicolosi et al. (1989) determined that a signal-to-noise ratio greater than 6 dB is needed for satisfactory communication. Later research involving the effect of noise began to use a multi-talker background rather than white noise more representative of everyday listening situations (i.e., *SPIN Test)*. And finally, more recent research by Buus et al. (1986) involving the effect of noise on non-native listeners compared to native listeners of English found that native listeners were able to tolerate a noise (white noise) level 12 dB greater than listeners with minimal exposure to English. Buus and colleagues suggested that this 12

dB difference could have the same effect as a 60 dB hearing loss relative to normal listeners.

The findings of the current study suggest that 6, 10 and 15 dB signal-to-noise ratios differ in their effect on speaker intelligibility. That is, regardless of accent, signal-to-noise ratio affected speaker intelligibility. A signal-to-noise ratio of 6 dB was the most difficult listening condition for all listeners, although for two speakers the differences between noise levels did not make a significant difference. For example, the moderate-strong speaker in the low predictability condition was difficult to understand. In this context, noise level did not seem to provide enough of a difference to improve listener accuracy (Figure 3-1). It appears that this speaker's accent was difficult to understand regardless of noise level. In fact, this speaker was difficult to understand in all conditions with the exception of the most optimal listening condition, 15 dB in the high predictability context. The opposite pattern was seen with the mild-moderate speaker. This speaker was so intelligible when linguistic predictability was high that noise level did not make a significant difference in listener accuracy. In the low predictability condition, noise affected this speaker when the signal-to-noise ratio decreased as is the general pattern (Figures 3-1 and 3-2).

The findings of this study would suggest that the conclusion of Nicolosi et al. (1989) that a signal-to-noise ratio greater than 6 dB is needed for satisfactory communication, may be too low for satisfactory communication in a fast-paced, high-risk listening environment such as air-traffic and emergency room communications as well as in many other environments. This level of noise (6 dB SNR) was inadequate for

even the native speaker when the linguistic context was of low predictability. The average listener accuracy with the native speaker in this difficult listening condition (6 dB SNR, LP) was 63% accurate (Table 4-1). This effect of noise (6 dB SNR) in both the high and low predictability contexts compromised communication when listeners were presented with the mild, moderate-strong, and strongest accents. At a 10 dB signal-to-noise ratio, listener accuracy dropped to less than 85% (practical significance level) in the low predictability context across all four non-native speakers. On the other hand, in the high-predictability context, listener accuracy decreased to a practical significance level only with the moderate-strong speaker. In the 15 dB noise condition all speakers were highly intelligible (>85%) in the high predictability context, but in the low predictability context only the native speaker remained highly intelligible.

In summary, the findings of this study confirm those of previous researchers that there is a difference in speaker intelligibility based on signal-to-noise ratio. The difference between the best listening condition (15 dB SNR, HP) in comparison to the worst listening condition (6 dB SNR, LP) definitely shows a significant difference in listener accuracy (Figures 3-3 and 3-5). In other words, listeners were more accurate at a 15 dB signal-to-noise ratio than they were at a signal-to-noise ratio of 6 dB. The differences were less obvious between speakers in the high predictability context where the native, mild, and mild-moderate speakers remained highly intelligible at all three noise levels.

The Effect of Linguistic Predictability

Several studies have compared the effect of linguistic predictability in noise on native and non-native listeners. For example, Florentine (1985) investigated the ability of non-native listeners and native listeners to take advantage of linguistic context in the presence of babble noise. Florentine's conclusions were supported by others (Bergman, 1980; Florentine et al., 1984; Nablelek & Donahue, 1984) who found that non-native speakers may demonstrate native-like speech recognition in quiet but have more difficulty understanding speech than native listeners in the presence of background noise. Florentine also concluded that in the presence of noise, non-native listeners did not benefit as much from contextual cues as did native listeners. These studies have focused on the comparison of non-native and native listeners listening to a standard American English speaker. The current study, on the other hand, examined the effect of predictability and noise when the speakers were non-native speakers of English and the listeners were native speakers of English.

The results of this study support the findings of Florentine and others in that native English listeners were able to use context to facilitate the comprehension of non-native speakers of English. For example, all listeners were more accurate in the high-predictability condition than they were in the low predictability condition (Table 4-1). These differences can also be seen in Figures 3-1 through 3-5. The differences in listener accuracy generally became greater as the signal-to-noise ratio decreased. Even the native speaker was more difficult for native listeners to comprehend when the signal-to-noise ratio was 6 dB in the low predictability condition, but they were not as affected by the

difference in predictability at the 10 and 15 dB signal-to-noise levels. The native speaker was highly intelligible at the higher signal-to-noise ratios, and listeners were less dependent upon context to help them identify the correct target word. As accent became stronger and noise levels became greater, the differences in listener accuracy between high and low predictability increased. This suggests that listeners were using context to help them determine the target word in difficult listening conditions, whereas in the low predictability condition listener responses were less accurate because context was not giving them additional cues toward the correct response. Florentine's (1985) conclusions that native speakers gain more from context than non-native speakers when listening to a native American speaker of English appear to apply also when native listeners are in a situation where they are listening to non-native speakers in a difficult listening environment. Florentine suggested that even highly proficient non-native listeners may lose as much as 30% of the information gathered by native listeners when the listening environment is compromised by noise. The findings of the current study support Florentine's conclusions. That is, native listeners were found to use linguistic predictability to help them identify target words when the listening environment was degraded by noise. As the noise conditions improved, listeners were less dependent upon context. These findings add to Florentine's conclusion by looking at it from the perspective of native listeners listening to non-native speakers. Therefore, the findings of this study seem to suggest that native listeners listening to non-native speakers in an equally compromised listening environment are likely to lose as much information as a non-native listener listening to a native speaker of English. For example, in air crew and

air traffic control communications, the speaker-listener dyad will change. When the listening environment is degraded by noise and the native speaker of English is the listener, he/she is likely to have as many difficulties understanding the non-native speaker as the non-native speaker would have listening to the native speaker.

In summary, the findings of this study suggest that there is a difference in speaker intelligibility based on linguistic context. That is, in all cases listeners were more accurate in the high-predictability condition than they were in the low-predictability condition. Listeners were able to take the most advantage of linguistic predictability when the listening condition was most compromised by noise (6 dB SNR). There was also a general trend that listeners were more reliant on context as accent became stronger in the better noise conditions (10 and 15 dB SNR).

It appears that there is a point where speech is so degraded by noise and/or strong accent that it is no longer intelligible with or without the benefit of contextual cues. Only speech that has a certain degree of overall intelligibility has the potential for further improvement with increased cues. Contextual cues, for example, may fail to upgrade the intelligibility of speech that is severely degraded (Sitler, Schiavetti, & Metz, 1983).

Definition and Measurement of Speech Intelligibility

Although it is often assumed by naive listeners that a particular degree of foreign accent will correspond to a particular level of intelligibility, Munro and Derwing (1995a) have offered evidence to the contrary. They concluded that foreign accent ratings did not predict intelligibility very well. One of the factors complicating the measurement of

intelligibility is the inconsistency in its definition. Smith and Nelson (1985) pointed out that the term intelligibility is often confused with the related terms comprehensibility and interpretability. It was intelligibility that was measured in its simplest sense, as the ability to identify a word or sentence, in this study.

This study was designed with the intent to measure intelligibility within the context of a sentence and within the context of realistic competing noise levels that might give a more accurate assessment of intelligibility than when words are tested in isolation and in an ideal listening environment. Because intelligibility is necessary for a message to be either comprehensible or interpretable, it was selected as the focus of measurement in this study.

Listeners were asked to identify the final word in each sentence. Intelligibility was then based on the number of correct responses in each condition. This was not a measure of comprehensibility because the listeners actual understanding of the meaning of the words or sentences was not examined. In fact, it was possible for the listeners to guess at the target word and still be correct. This would indicate that the speaker was intelligible but not necessarily comprehensible. However, the greater accuracy of listeners in the high predictability sentences would seem to indicate that they were comprehending the meaning to some extent. Interpretability, on the other hand, means that the speaker's intentions are understood. Such a judgment was outside the scope of this study as the speakers had no intention to actually convey a real message. The model sentences were provided for them to read. In fact, there is no evidence that the speakers actually understood the meaning of every word in every sentence.

The measurements of speech intelligibility selected for this study included a word-in-sentence identification task and listener ratings. The word identification task required the listener to write down the last word of a sentence produced by each speaker. The listener's responses were then compared to the target words to determine a percentage of speech intelligibility. The second method allowed the listener to rate the speaker's intelligibility based on the listener's overall impression of the speaker. Beukelman and Yorkston (1979) found a strong correlation between information transfer and word identification tests with dysarthric speakers. Beukelman and Yorkston (1980) also found that word identification tests were more sensitive and accurate, especially in the midrange of intelligibility, than were scaled scores of passages. Scaled scores often overestimate intelligibility. Although each method of measurement has strengths and weaknesses, Kent (1992) suggested that there may be no single test of intelligibility that can satisfy the research and clinical needs, therefore it is often appropriate and necessary to use several tools in the assessment of intelligibility. The two methods selected for this study were used to compare overall impressions as perceived by listeners with a more specific measurement of word in sentence identification percent correct score.

The findings of this study are consistent with those of Samar and Metz (1988, as cited in Schiavetti, 1992) that scaled scores often overestimate intelligibility. That is, when listeners were asked to rate the speakers' intelligibility pre- and post-task, the ratings were always at the high end of the 5-point Likert Scale. However, when compared to the actual word identification in sentences, listeners demonstrated a great

deal of variability ranging from no correct responses to five out of seven possible correct responses for the two speakers with the strongest accent in the most difficult listening condition. Listener responses also ranged from five to seven responses correct in the best listening condition with these two speakers. These findings suggest that while listeners rated the intelligibility of the speakers relatively high, speaker intelligibility was often low when listening conditions were degraded.

Linguistic and Non-Linguistic Factors

Denes and Pinson (1993) have pointed out that both linguistic and non-linguistic factors may influence intelligibility. Such factors are numerous and in many cases impossible to control. This study was designed to examine the influence of the linguistic factors of sentence predictability and accent and the influence of the non-linguistic factor of distracting noise in the environment.

Linguistically, the intelligibility of a word or utterance is influenced by our expectations based on our knowledge of the language. This knowledge includes the rules of the grammar, our familiarity with the topic, and our familiarity with the speaker. The linguistic context provides a great deal of information that influences our perception of what we expect to hear. The degree to which we can use linguistic factors may be reduced when we listen to a non-native speaker. Listeners who are uncertain that the non-native speaker will "follow the rules" may make fewer assumptions that would help them to interpret messages. Because non-native speakers of English vary in degree of foreign accent and the number of differences from what we expect, in extreme cases such differences can accumulate to the point where the native listener can no longer take

advantage of the linguistic context of the sentence or familiarity of the topic. That is, for the non-native speaker vowels and consonants are often pronounced differently from what we would normally expect from a native speaker. The speaker may be unfamiliar with the correct pronunciation or even the meaning of the word. In addition, the intonation patterns (stress, rhythm, timing) may deviate from what we expect. The combined effect of these linguistic factors makes the speech of non-native speakers more difficult to process and understand. There is evidence in the literature (Munro & Derwing 1995b; Schmid & Yeni-Komshian, 1999) that listeners do require more time to evaluate utterances produced by non-native speakers when compared to native speakers. The time factor was not explored in this study. However, it might be hypothesized that if the listeners were allowed more time to respond, their accuracy would increase.

It has been concluded by others (i.e., Buus et al., 1986; Catford, 1950; Gass & Varonis, 1984) that familiarity plays an important role in speaker intelligibility. Gass and Varonis found that familiarity with topic, familiarity with non-native speakers in general, and familiarity with a particular accent as well as with a particular speaker have an affect on intelligibility. In this study an effect of familiarity on intelligibility was not detected in the comparison of pre- and post-task ratings. However, this appeared to be a result of all speakers being rated as highly intelligible, even in the pre-rating task. In many situations, a particular communication environment (e.g., air-traffic and air crew communications, emergency room settings, classroom environments, telephone or radio communications, or even drive-through windows at fast food chains) presents a

situation where listeners are familiar with the topic. These settings generally have a limited context that leaves communication fairly predictable. However, when communication is vitally important, and if the language varies even slightly, a communication breakdown may occur. Such communication breakdowns could result in airplane crashes, medical errors, misunderstood directions, or food that was not ordered.

This study shows that native listeners of English were clearly affected by the linguistic factors of predictability and accent as well as by the third non-linguistic factor of noise (discussed above). Intelligibility was poorer when predictability was low, when perceived accent was greater, and when the signal-to-noise level was lowest. Listener responses were even affected by the mildest degree of a non-native accent especially in the most difficult listening condition. These findings suggest that as listening conditions became degraded by noise or linguistic predictability, listeners were less able to accommodate for foreign accent. On the other hand, as the listening conditions improved, listeners were more able to take advantage of the linguistic cues offered within the sentence to correctly identify the target word, even for the speaker with the strongest accent.

Practical Significance Level

Buus et al. (1986) investigated the effect of noise on non-native listeners compared to native listeners of English. Buus and colleagues determined that the amount of noise listeners could tolerate and still correctly repeat 50% of the sentences defined a Noise Tolerance Level. They found that native listeners were able to tolerate a greater noise level than non-native listeners with minimal exposure to English. Other

studies (using white noise) that were not concerned with foreign accent or context set a 50% criterion. Denes and Pinson (1993) stated that "normal conversation" can generally occur without much difficulty at a level where a 50% word articulation score can be achieved. In this study, listener tolerance levels have been presented in the context of a "practical significance level" when listener accuracy is less than 85%. This level was reached when the group of listeners correctly identified six out of seven words (85.7% referred to as 85% for ease of interpretation). The interaction between the three factors of interest in this study make it impossible to clearly designate one noise level or one accent level that would adversely impact listener accuracy at a practical level. However, noise is definitely an important factor with a 6 dB signal-to-noise ratio compromising the intelligibility of all the speakers, including the native speaker when the context was of low predictability (Table 4-1). When the native speaker and the two speakers with the mildest accents were heard under the same signal-to-noise ratios in a high predictability context, intelligibility was not compromised. So, even with the native speaker a noise tolerance level cannot be established independent of context. However, the intelligibility of two of the five speakers was compromised even at the high predictability level. It seems reasonable to conclude that a signal-to-noise ratio of 6 dB is not satisfactory for adequate communication between native and most non-native speakers of English regardless of linguistic context. A signal-to-noise ratio of 10 dB is also less than satisfactory for sufficient intelligibility for all accented speakers in a low predictability context. On the other hand, a signal-to-noise ratio of 10 dB in a high predictability context only compromised the speech intelligibility of one of the speakers

with the strongest accents. Therefore, in a communication setting where topic familiarity and listener familiarity are well established, a 10 dB signal-to-noise ratio may not cause difficulty. However, where the communication is critical to the lives of other people, the slightest deviation from what is linguistically expected may lead to disaster. In ideal communication situations where the noise level was low (i.e., 15 dB SNR) and the predictability was high, all speakers were highly intelligible. Degrading the situation by context alone reduced the intelligibility to below the practical significance level with the three speakers with the strongest accent.

In summary, the factors of interest in this study interact in such a way that a clear definition of noise level or degree of foreign accent cannot be defined. The findings suggest that when communication strays from what is expected linguistically, communication can be compromised even in the best noise condition (15 dB SNR). It is possible that we would find similar conclusions with the stronger accents when noise is not a factor at all.

Comparison with Pilot Study

Conclusions from the pilot study were based on a descriptive analysis versus a statistical analysis. Only six listeners were used in the pilot study, whereas, the current study analyzed the responses of 50 listeners. The number of stimuli was also increased from the 120 items in the pilot study to 210 items. This increased the number of responses from four to seven opportunities for each listener in each of the 30 conditions. The findings of both the pilot study and the current study were in agreement that all factors of accent, noise, and predictability contributed to listener

accuracy. Conclusions from the pilot study were that listener accuracy dropped significantly as accent strengthened and that the variability in accuracy among listeners increased as accent increased. That is, the range of accurate responses among the six listeners was very narrow with the native and mild speakers and wider with stronger accented speakers. Some listeners were accurate while other listeners had a great deal of difficulty with the speakers with the stronger accents. This trend was not seen in the current study when all 50 listeners as a group were examined across degrees of accent. In both the pilot study and the current study, listener accuracy improved as the noise condition improved. The pilot study also found that differences in listener accuracy between the two levels of predictability were as great as 48-50% in the two strongest accents. This finding was even stronger in the current study. In all cases, listener accuracy improved in the high predictability condition when compared to the low predictability condition.

The pilot study and the current study concur in that listeners rated all speakers as having high intelligibility regardless of degree of accent. Listeners did not begin to lower their intelligibility rating of the speakers until they heard the two speakers with the strongest accent. The data from the current study shows that only two to four of the 50 listeners rated the mild and mild-moderate speakers with an intelligibility level of 4 on the 5-point Likert scale, only one to two listeners rated the two strongest speakers with a 3, and one listener rated the strongest speaker with a 2. A rating of 5 indicated the listener understood at least 98-100% of the entire message, whereas a rating of 2 indicated the listener had a great deal of difficulty understanding the speaker, only

identifying single words or a few phrases. Despite the lower ratings with the stronger accents, almost all listeners (94-96%) still rated the two speakers with the strongest accents at a level of at least 4, indicating that they understood most of the speech sample with the exception of a few words or phrases (Appendix H).

In summary, the findings of the current study are in agreement with those found in the pilot study with the support of statistical analysis. Because the trends were similar in both studies it may be more useful to use a smaller number of listeners who can reasonably be compared individually for a more qualitative analysis that sometimes gets lost with large amounts of data. However, the greater number of opportunities for listeners to hear all conditions added strength to the inferences drawn from the statistical conclusions.

Generalization of Findings

Several researchers have raised the issue of our ability to generalize laboratory research findings from artificial listening situations to listening in real-world situations (i.e., Gagne, 1994; Kalikow et al., 1977; Tyler, 1994). The listening conditions in this study were designed to determine the effects of the controlled factors--accent, noise, and linguistic predictability--in a setting representative of a realistic listening environment. The manipulation of accent, noise, and linguistic predictability were intended to examine a range of speakers in a variety of conditions that could possibly be generalized to other listening environments. In many situations listeners also would have the benefit of visual information that can facilitate the comprehension of speech that may be otherwise difficult to comprehend. Such use of multiple modalities was not explored here.

However, the study was designed to represent communication settings where visual information is not available to the listener such as radio or telephone communication. Multi-talker babble noise was selected as the background noise for this study because it represents a more difficult listening condition than a steady-state noise for which listeners can often accommodate.

Although the setting in this study was experimental and conditions were controlled, it was designed to have a broad application to several listening conditions. Therefore, the sentences chosen were not specific to one particular setting. The conditions in this study may not exactly match any particular setting; however, the noise conditions were representative of realistic levels of background noise where communication often takes place. The non-native speakers were narrowly selected to represent one language, and they represented the high-end of the scale in language proficiency. However, it does seem likely that these findings can be generalized to some degree to the intelligibility of other non-native speakers frequently encountered in international communication situations. If anything, these findings are conservative if we were to compare non-native speakers who are less proficient speakers of English or who may have a greater number of phonological differences (i.e., tonal languages) in their native language that carry over into their production of English.

Strengths and Limitations

Strengths

This study was designed to investigate the impact of foreign accent, noise, and linguistic predictability on the intelligibility of non-native speakers of English. Other

studies (Buus et al., 1986; Florentine, 1985; Mayo et al., 1997) have examined the effect of noise and linguistic predictability on the perception of non-native listeners of English rather than the effect of foreign accent in conditions of noise and varying predictability on speaker intelligibility. This study focused on speaker intelligibility as a function of degree of accent. It is now a common occurrence for English to be spoken by non-native speakers in many realms of business and daily living. Even when communication takes place between native speakers of English, it is often compromised when listening conditions become degraded by noise or poor transmission systems. Communication becomes further compromised when the same information is relayed by a speaker with a foreign accent. This study was designed to identify the point at which accent begins to interfere with communication in these adverse listening conditions. It is a preliminary step in examination of these communication situations by measuring listener responses at fixed levels of accent, noise, and predictability. At this time there is very little research that looks at the combination of these factors in such detail. The findings of this study contribute to the literature by offering results that indicate how intricately these factors are interwoven in the communication process.

Limitations

The findings of this study hold for male, Brazilian-Portuguese speakers. Caution should be applied to generalization of the findings to other languages and to female speakers. The findings are also relevant to listeners who had no prior experience with non-native speakers of Brazilian-Portuguese. It is quite likely that familiarity with the accent could facilitate comprehension.

The study was designed using two randomized versions of the sentences and speakers. The experimental stimuli were also presented to the listeners as a group for the purposes of efficiency in test administration and data collection. A better design would be for each listener to have heard a completely new randomized scheme of the sentences. In this way responses would not be as affected by listening to the better speakers first. For example, in some cases listeners heard the native speaker or mild speaker before they heard the other speakers producing the same sentences. The opportunity to hear the sentences produced by the native and/or mild accented speakers first could influence subsequent listener responses to the speakers with the stronger accents that occurred later in the randomized list. To accommodate for this weakness or limitation, a computerized version could be designed where the sentences could be randomized for each individual listener.

Suggestions for Further Research

This research not only contributes to the understanding of speech perception under these adverse listening conditions, but it also has several practical applications. For example, the information can be used to contribute to development of accent reduction programs. Further acoustical analysis of the actual responses may provide information on patterns of phonemes, intonation, and prosody that can interfere with understanding. Acoustical analysis can also help determine if there were patterns of words or particular features where noise was a specific problem. It would be interesting to investigate what acoustic features made one speaker intelligible, leaving another speaker unintelligible. Studies such as this should facilitate the choice of methods used

for determining specific types of training programs and aspects that will be most effective in facilitating the perception and production of English by non-native speakers.

Although the sample of listeners were quite large, 50, the number of non-native speakers was small representing only a limited sample of degree of accent. Only one non-native language was represented. Only speakers who were determined to be highly proficient in English were used in this study. It would be useful to replicate this study with speakers from other languages and with a broader range of English proficiency. Because communication in international settings is often an exchange between a non-native speaker of one language and a non-native speaker of another language both speaking a third language, it would also be useful to replicate this study with a group of non-native listeners. It is speculated that non-native listeners of another language would have more difficulty understanding the same sentences because they may be less able to take advantage of the linguistic and non-linguistic cues expected by native listeners.

This study was designed to have a broad application to several listening conditions. Therefore, the sentences chosen were not specific to one particular setting. In a future study it would be useful to design sentences that would take into consideration the effect of topic familiarity. For example, sentences specific to air-traffic control and air-crew communications situations or emergency room situations may provide more information than the general sentences used in this study.

Summary and Conclusions

It is clear that all three factors examined in this study have a role in the intelligibility of non-native speakers of English. For many years researchers have investigated the effect of noise on speaker intelligibility (e.g., Denes & Pinson, 1993; Nicolosi et al., 1989), the effect of noise and linguistic predictability on speaker intelligibility (e.g., Kalikow et al., 1977), and the effect of accent on speaker intelligibility (e.g., Munro & Derwing, 1995a, 1995b; Schmid & Yeni-Komshian, 1999). Some studies have even combined these factors to examine the effects of noise and varying linguistic contexts in the perception of non-native speakers of English when listening to native speakers of English (e.g., Florentine, 1985; Mayo et al., 1997). However, studies have not examined the effect of accent on speaker intelligibility in conditions of noise and linguistic predictability in environments representative of real-life, day-to-day communication. The current study suggests that although noise and linguistic predictability do affect speaker intelligibility, the degree of accent can further compromise listener accuracy in a word-in-sentence identification task. The findings of this study offer an answer to the question proposed by Munro and Derwing (1995b).

> It is not known whether the effects of noise or filtering (e.g., in telephone or radio transmission, in noisy rooms, or at variable loudness levels) have the same degree of impact on the processing time or comprehensibility of accented speech as on native-produced speech, or whether the effects of such conditions vary as a function of degree of accent. (p. 303)

In fact, the effects of noise and linguistic predictability do appear to vary as a function of degree of accent.

APPENDIX A
LISTENER QUESTIONNAIRE
STUDY OF NON-NATIVE SPEAKERS OF ENGLISH

Name________________________ Age_____________

Native Language_______________ Gender: Male or Female
(circle one)

Other languages spoken____
(fluently)

Hearing Screening: _________ Tape #___________

1. Indicate the degree of exposure you have had with speakers with a foreign accent

1	2	3	4	5
(very limited exhanges)				(everyday/extended conversations)

2. List the languages you have had exposure to______________________________

Based on the scale above, please give a rating for each language you listed in number 2. Place that rating next to each language you listed.

3. Have you had experience listening to speakers of Brazilian Portuguese? Yes or No

If yes, please indicate your degree of exposure

1	2	3	4	5
(very limited exhanges)				(everyday/extended

conversations)

5. Have you had any formal musical training? Yes or No

If yes, how many years _________ what instrument(s) ______________________

APPENDIX B
GRIFFITHS INTELLIGIBILITY TEST

LISTENER ID____________________ **DATE**____________________

LIST: A

1	2	3	4	5
BAT	DIG	LASH	PICK	SHEEN
BATCH	DIN	LACK	PIT	SHEAVE
BASH	DID	LASS	PIP	SHEATHE
BASS	DIM	LAUGH	PIG	SHEATH
BADGE	DILL	LATH	PITCH	SHEAF

6	7	8	9	10
LAWS	DONE	MAT	PUP	SING
LONG	DUD	MAD	PUFF	SIP
LOG	DUNG	MATH	PUB	SIN
LODGE	DUB	MAN	PUCK	SIT
LOB	DUG	MASS	PUS	SICK

11	12	13	14	15
WIG	FILL	BEIGE	HATH	SUD
WITH	FIG	BASE	HASH	SUM
WIT	FIN	BAYED	HALF	SUB
WITCH	FIZZ	BATHE	HAVE	SUN
WICK	FIB	BAYS	HAS	SUNG

16	17	18	19	20
DUMB	LEAVE	PASS	WE'RE	TAB
DUB	LEIGE	PATH	WHEEL	TAN
DOTH	LEACH	PACK	WEAVE	TAM
DUFF	LEASH	PAD	WEED	TANG
DOVE	LEAD	PAT	WEAN	TAP

21	22	23	24	25
CUFF	TOSS	PEAK	SAD	TEETHE
CUB	TAJ	PEAS	SAT	TEAR
CUT	TONG	PEAL	SAG	TEASE
CUP	TALKS	PEACE	SACK	TEEL
CUD	TOG	PEAT	SAP	TEETH

26	27	28	29	30
LED	FIN	FEEL	RENT	ZIP
SHED	TIN	REEL	BENT	LIP
RED	SHIN	SEAL	WENT	NIP
WED	KIN	ZEAL	DENT	GYP
FED	THIN	VEAL	TENT	SHIP

31	32	33	34	35
SOLD	BARK	TEN	HIP	NEST
COLD	DARK	PEN	RIP	BEST
HOLD	MARK	DEN	TIP	VEST
TOLD	LARK	HEN	DIP	REST
GOLD	PARK	THEN	LIP	WEST

36	37	38	39	40
DIG	GALE	PIN	TOP	BUST
BIG	PALE	SIN	HOP	JUST
WIG	TALE	TIN	POP	RUST
RIG	BALE	WIN	COP	GUST
PIG	MALE	FIN	SHOP	DUST

41	42	43	44	45
KICK	PEEL	THIN	YORE	MAT
CHICK	FEEL	TIN	GORE	VAT
THICK	EEL	CHIN	WORE	THAT
PICK	HEEL	SHIN	LORE	FAT
SICK	KEEL	GIN	ROAR	RAT

46	47	48	49	50
WILL	SHAME	THEE	VIE	WAY
HILL	GAME	DEE	THY	MAY
KILL	CAME	LEE	FIE	GAY
TILL	SAME	KNEE	THIGH	THEY
BILL	TAME	ZEE	HIGH	NAY

APPENDIX C
SPIN SENTENCE LISTS 2.1 - 2.8

SPIN TEST SENTENCE LIST 2.1

1. The watchdog gave a warning growl.
2. She made the bed with clean sheets.
3. The old man discussed the dive.
4. Bob heard Paul called about the strips.
5. I should have considered the map.
6. The old train was powered by steam.
7. He caught the fish in his net.
8. Miss Brown shouldn't discuss the sand.
9. Close the window to stop the draft.
10. My T.V. has a twelve-inch screen.
11. They might have considered the hive.
12. David has discussed the dent.
13. The sandal has a broken strap.
14. The boat sailed along the coast.
15. Crocodiles live in muddy swamps.
16. He can't consider the crib.
17. The farmer harvested his crop.
18. All the flowers were in bloom.
19. I am thinking about the knife.
20. David does not discuss the hug.
21. She wore a feather in her cap.
22. We've been discussing the crates.
23. Miss black knew about the doll.
24. The Admiral commands the fleet.
25. She couldn't discuss the pine.
26. Miss Black thought about the lap.
27. The beer drinkers raised their mugs.
28. He was hit by a poisoned dart.
29. The bread was made from whole wheat.
30. Mr. Black knew about the pad.
31. You heard Jane called about the van.
32. I made the phone call from a booth.
33. Tom wants to know about the cake.
34. She's spoken about the bomb.
35. The cut on his knee formed a scab.
36. We hear you called about the lock.
37. The old man discussed the yell.
38. His boss made him work like a slave.
39. The farmer baled the hay.
40. They're glad we heard about the track.
41. A termite looks like an ant.
42. Air mail requires a special stamp.
43. Football is a dangerous sport.
44. Sue was interested in the bruise.
45. Ruth will consider the herd.
46. We saw a flock of wild geese.
47. The girl talked about the gin.
48. Paul can't discuss the wax.
49. Drop the coin through the slot.
50. I hope Paul asked about the mate.

SPIN TEST SENTENCE LIST 2.2

1. You're glad they heard about the slave.
2. The girl knows about the swamps.
3. Hold the baby on your lap.
4. For your birthday I baked a cake.
5. The railroad train ran off the track.
6. They did not discuss the screen.
7. They were interested in the strap.
8. Tear off some paper from the pad.
9. I had a problem with the bloom.
10. Peter should speak about the mugs.
11. The fruit was shipped in wooden crates.
12. The rancher rounded up his heard.
13. She wants to speak about the ant.
14. We're discussing the sheets.
15. The boy would discuss the scab.
16. The lonely bird searched for its mate.
17. Tom could have thought about the sport.
18. You'd been considering the geese.
19. They drank a whole bottle of gin.
20. One ht beach we play in the sand.
21. Mr. Black considered the fleet.
22. The airplane went into a dive.
23. We're lost so let's look at the map.
24. I want to know about the crop.
25. Household goods are moved in a van.
26. The honey bees swarmed round the hive.
27. Betty has talked about the draft.
28. Tom discussed the hay.
29. Jane was interested in the stamp.
30. The airplane dropped a bomb.
31. Cut the bacon into strips.
32. I had not thought about the growl.
33. The drowning man let out a yell.
34. I gave her a kiss and a hug.
35. Paul should know about the net.
36. I cut my finger with a knife.

37. The candle flame melted the wax.
38. Tom heard Jane called about the booth.
39. We can't consider the wheat.
40. This key won't fit in the lock.
41. We have not discussed the steam.
42. Miss Brown might consider the coast.
43. Mr. Brown can't discuss the slot.
44. The little girl cuddled her doll.
45. Tom fell down and got a bad bruise.
46. He hasn't considered the dart.
47. The furniture was made of pine.
48. How did your car get that dent?
49. Mr. Smith thinks about the cap.
50. The baby slept in his crib.

SPIN TEST SENTENCE LIST 2.3

1. A rose bush has prickly thorns.
2. We should have considered the juice.
3. The shipwrecked sailors built a raft.
4. Bob could have known about the spoon.
5. Ruth poured the water down the drain.
6. The boy gave the football a kick.
7. Bill might discuss the foam.
8. The cop wore a bullet-proof vest.
9. Tom could not discuss the barn.
10. You were considering the gang.
11. After his bath, he wore a robe.
12. Nancy should consider the fist.
13. I can't guess so give me a hint.
14. The soup was served in a bowl.
15. I've spoken about the pile.
16. Jane has a problem with the coin.
17. The bomb exploded with a blast.
18. Mary could not discuss the tack.
19. They have a problem with the limb.
20. Nancy had considered the sleeves.
21. Lubricate the car with grease.
22. The workers are digging a ditch.
23. Bill heard Tom called about he coach.
24. They marched to the beat of the drum.
25. No one was injured in the crash.
26. The old man thinks about the mast.
27. The sailor swabbed the deck.
28. Tom will discuss the swan.
29. Ann was interested in the breath.
30. his nozzle sprays a fine mist.
31. Ruth hopes he heard about the hips.
32. Tom is talking about the fee.
33. Miss Smith considered the scare.
34. The ship's Captain summoned his crew.
35. They fished in the babbling brook.
36. The hockey player scored a goal.
37. David should consider the blame.
38. They played a game of cat and mouse.
39. He's glad you called about the jar.
40. Tom will discuss the cot.
41. The steamship left on a cruise.
42. She faced them with a foolish grin.
43. He hopes Tom asked about the bar.
44. Miss Black could have discussed the rope.
45. A chimpanzee is an ape.
46. He wiped the sink with a sponge.
47. We shipped the furniture by truck.
48. Ruth's Grandmother discussed the broom.
49. I've been considering the crown.
50. A bear has a thick coat of fur.

SPIN TEST SENTENCE LIST 2.4

1. I want to speak about the crash.
2. Harry slept on the folding cot.
3. She's glad Jane asked about the drain.
4. The doctor charged a low fee.
5. He had considered the robe.
6. I haven't discussed the sponge.
7. The guilty one should take the blame.
8. You cannot have discussed the grease.
9. The cookies were kept in a jar.
10. Let's invite the whole gang.
11. Mr. White discussed the cruise.
12. The sport shirt has short sleeves
13. They knew about the fur.
14. We've spoken about the truck.
15. The cushion was filled with foam.
16. How long can you hold your breath?
17. She wants to talk about the crew.
18. The cow was milked in the barn.
19. The accident gave me a scare.
20. The kitten climbed out on a limb.
21. You're glad she called about the bowl.
22. The man could not discuss the mouse.
23. He tossed the drowning man a rope.
24. You hope they asked about the vest.
25. You want to talk about the ditch.
26. Stir your coffee with a spoon.
27. We hear she called about the drum.
28. Bob stood with his hands on his hips.
29. The teacher sat on a sharp tack.
30. She might have discussed the ape.
31. The storm broke the sailboat's mast.
32. At breakfast he drank some juice.
33. He hit me with a clenched fist.
34. Peter knows about the raft.
35. The old man considered the kick.
36. We have not thought about the hint.
37. The team was trained by their coach.
38. Bill hopes Paul heard about the mist.
39. The king wore a golden crown.
40. The sand was heaped in a pile.
41. The boy can't talk about he thorns.
42. Miss Brown will speak about the grin.
43. The duck swam with the white swan.
44. Let's decide by tossing a coin.

45. She has a problem with the goal.
46. Jane didn't think about the brook.
47. He hears she asked about the deck.
48. He got drunk in the local bar.
49. The girl swept the floor with a broom.
50. The class will consider the blast.

SPIN TEST SENTENCE LIST 2.5

1. Miss White would consider the mold.
2. Ruth has a problem with the joints.
3. The boy might consider the trap.
4. To store his wood he built a shed.
5. The lion gave an angry roar.
6. He is considering the throat.
7. They hope he heard about the rent.
8. The car was parked at the curb.
9. Peter should consider the bow. (as in "no")
10. The old woman discussed the thief.
11. A round hole won't take a square peg.
12. You're discussing the plot.
13. The woman knew about the lid.
14. Peter dropped in for a brief chat.
15. You were interested in the scream.
16. The gambler lost the bet.
17. The burglar escaped with the loot.
18. He could discuss the bread.
19. He was scared out of his wits.
20. He doesn't discuss the mop.
21. Even was made from Adam's rib.
22. Get the bread and cut me a slice.
23. Bill won't consider the brat.
24. We heard the ticking of the clock.
25. Greet the heroes with loud cheers.
26. This camera is out of film.
27. Ruth wants to speak about the sling.
28. My jaw aches when I chew gum.
29. The man could consider the spool.
30. The bloodhound followed the trail.
31. The doctor prescribed the drug.
32. He rode off in a cloud of dust.
33. He was interested in the hedge.
34. Ruth hopes she called about the junk.
35. Playing checkers can be fun.
36. We're glad Ann asked about the fudge.
37. The super highway has six lanes.
38. Unlock the door and turn the knob.
39. Ruth is speaking about the meal.
40. Maple syrup is made from sap.
41. Bill cannot consider the den.
42. We are speaking about the prize.
43. The car drove off the steep cliff.
44. Miss Smith couldn't discuss the row. (as in "no")
45. The glass had a chip on the rim.
46. Old metal cans were made with tin.
47. Miss White thinks about the tea.
48. Miss White doesn't discuss the cramp.
49. That job was an easy task.
50. Mr. White spoke about the firm.

SPIN TEST SENTENCE LIST 2.6

1. Throw out all this useless junk.
2. She cooked him a hearty meal.
3. Her entry should win first prize.
4. Ruth could have discussed the wits.
5. We could discuss the dust.
6. The stale bread was covered with mold.
7. The firemen heard her frightened scream.
8. We spoke about the knob.
9. Your knees and your elbows are joints.
10. I ate a piece of chocolate fudge.
11. Paul hopes we heard about the loot.
12. Instead of a fence, plant a hedge.
13. The story had a clever plot.
14. David might consider the fun.
15. The landlord raised the rent.
16. Paul could not consider the rim.
17. He heard they called about the lanes.
18. Her hair was tied with a blue bow. (as in "no")
19. They had a problem with the cliff.
20. He's employed by a large firm.
21. Harry will consider the trail.
22. We are considering the cheers.
23. To open the jar, twist the lid.
24. She has known about the drug.
25. Bill had a problem with the chat.
26. We hear they asked about the shed.
27. The swimmer's leg to a bad cramp.
28. Jane had not considered the film.
29. Our seats wee in the he second row. (as in "no)
30. Jane did not speak about the slice.
31. Paul was interested in the sap.
32. I am discussing the task.
33. The thread was wound on a spool.
34. They tracked the lion to his den.
35. Ruth has discussed the peg.
36. Spread some butter on your bread.
37. Tom is considering the clock.
38. He's thinking about the roar.
39. A spoiled child is a brat.
40. I should have known about the gum.
41. Keep your broken arm in a sling.
42. The mouse was caught in the trap.
43. They heard I asked about the bet.
44. I've got a cold and a sore throat.
45. Betty doesn't discuss the curb.
46. He had a problem with the tin.
47. Ruth poured herself a cup of tea.
48. The house was robbed by a thief.
49. He wants to know about the rib.
50. Wash the floor with a mop.

SPIN TEST SENTENCE LIST 2.7

1. I did not know about the chunks.
2. The chicken pecked the corn with its beak.
3. Bob could consider the pole.
4. The judge is sitting on the bench.
5. Mr. Smith knew about the bay.
6. You've considered the seeds.
7. The heavy rains caused a flood.
8. For dessert he had apple pie.
9. She hopes Jane called about the calf.
10. The detectives searched for a clue.
11. Mary hasn't discussed the blade.
12. The chicks followed the mother hen.
13. Mr. Brown thinks about the vault.
14. Bob was considering the clerk.
15. We camped out in our tent.
16. Paul took a bath in the tub.
17. Mary can't consider the tide.
18. The old man talked about the lungs.
19. The candle burned with a bright flame.
20. My son has a dog for a pet.
21. Bob has discussed the splash.
22. The plow was pulled by an ox.
23. The flood took a heavy toll.
24. Mr. Smith spoke about the aid.
25. Mary had considered the spray.
26. The pond was full of croaking frogs.
27. The girl should not discuss the gown.
28. Please wipe your feet on the mat.
29. Ruth hopes Bill called about the cop.
30. We will consider the debt.
31. Peter could consider the dove.
32. She shortened the hem of her skirt.
33. The cabin was made of logs.
34. Bill can't have considered the wheels.
35. He has a problem with the oath.
36. The dealer shuffled the cards.
37. The shepherd watched his flock of sheep.
38. The flashlight casts a bright beam.
39. We could consider the feast.
40. The scarf was made of shiny silk.
41. The guests were welcomed by the host.
42. Betty has considered the bark.
43. The sick child swallowed the pill.
44. Paul should have discussed the flock.
45. Tighten the belt by a notch.
46. She might discuss the crumbs.
47. Tom has not considered the glue.
48. The swimmer dove into the pool.
49. Tom has been discussing the beads.
50. Follow this road around the bend.

SPIN TEST SENTENCE LIST 2.8

1. The bird of peace is the dove.
2. Tom had spoken about the pill.
3. The cigarette smoke filled his lungs.
4. They've considered the sheep.
5. Cut the meat into small chunks.
6. Watermelons have lots of seeds.
7. The man should discuss the ox.
8. Miss Smith knows about the tub.
9. Raise the flag up the pole.
10. Peter has considered the mat.
11. The bride wore a white gown.
12. She might consider the pool.
13. We swam at the beach at high tide.
14. The poor man was deeply in debt.
15. She's glad Bill called about the beak.
16. Harry had thought about the logs.
17. Banks keep their money in a vault.
18. The witness took a solemn oath.
19. Bill didn't discuss the hen.
20. Ruth must have known about the pie.
21. The shepherds guarded their flock.
22. Bob has considered the tent.
23. We're speaking about the toll.
24. A bicycle has two wheels.
25. Ann works in the bank as a clerk.
26. Tom won't consider the silk.
27. Ruth had a necklace of glass beads.
28. She's discussing the beam.
29. Paul hit the water with a splash.
30. The nurse gave him first aid.
31. The wedding banquet was a feast.
32. Nancy didn't discuss the skirt.
33. The girl should consider the flame.
34. Tree trunks are covered with bark.
35. Break the dry bead into crumbs.
36. Mr. Black has discussed the cards.
37. The woman considered the notch.
38. The man spoke about the clue.
39. The boat sailed across the bay.
40. I'm talking about the bench.
41. They heard I called about the pet.
42. The cow gave birth to a calf.
43. I'm glad you heard about the bend.
44. It was stuck together with glue.
45. The woman talked about the frogs.
46. Bob was cut by the jackknife's blade.
47. Paul was arrested by the cops.
48. Bill heard we asked about the host.
49. Kill the bugs with this spray.
50. The class should consider the flood.

APPENDIX D
SPIN SENTENCES USED FOR THE LISTENING TASK

Training Sentences: from List 2.8* (20 dB SNR)
1. The bride wore a white *gown.* (HP)
2. Banks keep their money in a *vault.* (HP)
3. She's glad Bill called about the *beak.* (LP)
4. The nurse gave him first *aid.* (HP)
5. The boat sailed across the *bay.* (HP)

*one sentence from each speaker

SPIN Sentences Randomly Selected For The Listening Task**

<u>List 2.1</u> (6 dB SNR)

LP	HP
1. They might have considered the hive.	1. The cut on his knee formed a scab.
2. The old man discussed the dive.	2. Air mail requires a special stamp.
3. She's spoken about the bomb.	3. She made the bed with clean sheets.
4. The girl talked about the gin.	4. He was hit by a poisoned dart.
5. David does not discuss the hug.	5. The watchdog gave a warning growl.
6. Miss Black thought about the lap.	6. She wore a feather in her cap.
7. I should have considered the map.	7. He caught the fish in his net.

<u>List 2.3</u> (10 dB SNR)

LP	HP
1. Nancy had considered the sleeves.	1. After his bath, he wore a robe.
2. The old man thinks about the mast.	2. The shipwrecked sailors built a raft.
3. We should have considered the juice.	3. They fished in the balling brook.
4. Tom will discuss the swan.	4. The ship's Captain summoned his crew.
5. Bill might discuss the foam.	5. We shipped the furniture by truck.
6. Bill heard Tom called about the coach.	6. She faced them with a foolish grin.
7. Ruth hopes he heard about the hips.	7. Ruth poured the water down the drain.

<u>List 2.6</u> (15 dB SNR)

LP	HP
1. Paul hopes we heard about the loot.	1. Her entry should win first prize.
2. Ruth has discussed the peg.	2. They tracked the lion to his den.
3. We hear they asked about the shed.	3. Ruth poured herself a cup of tea.
4. We could discuss the dust.	4. He's employed by a large firm.
5. Jane had not considered the film.	5. The mouse was caught in the trap.
6. Harry will not consider the trail.	6. She cooked him a hearty meal.
7. He's thinking about the roar.	7. Spread some butter on your bread.

** Each speaker produced all sentences. All sentences were randomized for two separate audio-recordings. (Tapes A and B).

APPENDIX E
LISTENER INSTRUCTIONS

Listener Instructions:

After listening to the pre-recorded tape of the non-native speaker please circle the appropriate response.

Accent: Phonetic traits of an individual's native language carried over into a second, foreign language (Nicolosi, Harryman & Kresheck, 1989).

What **degree of accent** would you assign this speaker?

No detectable foreign accent	**Mild**	**Moderate**	**Strong**
0	**1 2 3**	**4 5 6**	**7 8 9**

Intelligibility: word/utterance recognition; a word/utterance is considered to be unintelligible when the listener is unable to make it out and, thus, to repeat (or write) it. (Smith & Nelson, 1985)

Overall, how would you rate the **intelligibility level** of this speaker, using the 5-point scale provided?

1 = I did not understand the speaker at all
2 = I had a lot of difficulty understanding the speaker--I could only pick out single words or a few phrases.
3 = I was able to understand about 50% of the speech sample.
4 = I understood most of the speech sample with the exception of a few words or phrases.
5 = I understood at least 98-100% of the entire message.

Scale adapted from the National Technical Institute for the Deaf (NTID) Rating Scale (Subtelny, Orlando & Whitehead, 1981)

APPENDIX F
PRE- AND POST-RATING FORM

Rater ____________________________ **Pre or Post Rating** (Circle one)

1. Please circle the category which represents the **degree** of foreign accent you feel best describes each speaker.

2. Please circle the **intelligibility level** of each speaker, using the 5-point scale provided.

Speaker # 1

No detectable foreign accent	Mild	Moderate	Strong
0	1 2 3	4 5 6	7 8 9

Intelligibility Rating:
1 = I did not understand the speaker at all
2 = I had a lot of difficulty understanding the speaker--I could only pick out single words or a few phrases.
3 = I was able to understand about 50% of the speech sample.
4 = I understood most of the speech sample with the exception of a few words or phrases.
5 = I understood at least 98-100% of the entire message

Speaker # 3

No detectable foreign accent	Mild	Moderate	Strong
0	1 2 3	4 5 6	7 8 9

Intelligibility Rating:
1 = I did not understand the speaker at all
2 = I had a lot of difficulty understanding the speaker--I could only pick out single words or a few phrases.
3 = I was able to understand about 50% of the speech sample.
4 = I understood most of the speech sample with the exception of a few words or phrases.
5 = I understood at least 98-100% of the entire message

Speaker # 6

No detectable foreign accent	Mild			Moderate			Strong		
0	1	2	3	4	5	6	7	8	9

Intelligibility Rating:
1 = I did not understand the speaker at all
2 = I had a lot of difficulty understanding the speaker--I could only pick out single words or a few phrases.
3 = I was able to understand about 50% of the speech sample.
4 = I understood most of the speech sample with the exception of a few words or phrases.
5 = I understood at least 98-100% of the entire message

Speaker # 7

No detectable foreign accent	Mild			Moderate			Strong		
0	1	2	3	4	5	6	7	8	9

Intelligibility Rating:
1 = I did not understand the speaker at all
2 = I had a lot of difficulty understanding the speaker--I could only pick out single words or a few phrases.
3 = I was able to understand about 50% of the speech sample.
4 = I understood most of the speech sample with the exception of a few words or phrases.
5 = I understood at least 98-100% of the entire message

Speaker # 5

No detectable foreign accent	Mild			Moderate			Strong		
0	1	2	3	4	5	6	7	8	9

Intelligibility Rating:
1 = I did not understand the speaker at all
2 = I had a lot of difficulty understanding the speaker--I could only pick out single words or a few phrases.
3 = I was able to understand about 50% of the speech sample.
4 = I understood most of the speech sample with the exception of a few words or phrases.
5 = I understood at least 98-100% of the entire message

APPENDIX G
EXAMPLE OF LISTENER RESPONSE FORM

Listener Response Form __________ **Listener ID**______________

Training Items: Word Identification

1. __________________
2. __________________
3. __________________
4. __________________
5. __________________

Rater Code	Word Identification	Rater Code	Word Identification
________	1. ____________	________	26. ____________
________	2. ____________	________	27. ____________
________	3. ____________	________	28. ____________
________	4. ____________	________	29. ____________
________	5. ____________	________	30. ____________
________	6. ____________	________	31. ____________
________	7. ____________	________	32. ____________
________	8. ____________	________	33. ____________
________	9. ____________	________	34. ____________
________	10. ____________	________	35. ____________
________	11. ____________	________	36. ____________
________	12. ____________	________	37. ____________
________	13. ____________	________	38. ____________
________	14. ____________	________	39. ____________
________	15. ____________	________	40. ____________
________	16. ____________	________	41. ____________
________	17. ____________	________	42. ____________
________	18. ____________	________	43. ____________
________	19. ____________	________	44. ____________
________	20. ____________	________	45. ____________
________	21. ____________	________	46. ____________
________	22. ____________	________	47. ____________
________	23. ____________	________	48. ____________
________	24. ____________	________	49. ____________
________	25. ____________	________	50. ____________

APPENDIX H
PRE- AND POST-LISTENER RATINGS OF ACCENT AND INTELLIGIBILITY

		Accent Ratings (Scale: 0-9)					Intelligibility Ratings (Scale: 1-5)				
Lstnr ID	Pre/Post	N	M	MM	MS	S	N	M	MM	MS	S
1	Pre	0	2	3	4	4	5	5	5	5	4
2	Pre	0	1	3	3	3	5	5	5	5	5
3	Pre	0	2	2	6	5	5	5	5	4	4
4	Pre	0	3	5	5	7	5	5	5	4	3
5	Pre	0	2	6	5	7	5	5	5	5	4
6	Pre	0	2	3	4	7	5	5	5	5	4
7	Pre	0	3	6	7	7	5	5	5	5	5
8	Pre	0	6	5	7	5	5	5	5	4	4
9	Pre	0	3	5	3	5	5	5	5	5	4
10	Pre	0	2	6	4	6	5	5	5	5	5
11	Pre	0	2	3	3	4	5	5	5	5	5
12	Pre	0	2	5	6	7	5	5	5	5	4
13	Pre	0	3	4	4	6	5	4	4	4	4
14	Pre	0	3	3	5	4	5	5	5	5	4
15	Pre	0	1	2	5	5	5	5	5	4	4
16	Pre	0	3	5	6	5	5	5	5	5	5
17	Pre	0	3	4	8	7	5	5	5	5	5
18	Pre	0	3	6	7	6	5	5	5	5	2
19	Pre	0	3	6	5	6	5	5	4	3	3
20	Pre	0	2	4	5	3	5	5	5	5	4
21	Pre	0	1	2	2	3	5	5	5	5	5
22	Pre	0	2	4	5	4	5	5	5	5	4
23	Pre	0	3	6	6	7	5	5	5	5	5
24	Pre	0	5	6	7	7	5	5	5	4	5
25	Pre	0	2	3	2	6	5	5	5	5	4
26	Pre	0	1	4	5	6	5	5	5	4	4
27	Pre	0	1	3	4	6	5	5	5	5	5
28	Pre	0	3	5	6	5	5	5	5	5	5
29	Pre	0	1	2	5	3	5	5	5	5	4
30	Pre	0	2	5	7	5	5	5	5	5	5
31	Pre	0	1	4	2	7	5	5	5	5	4
32	Pre	0	1	1	2	4	5	5	5	5	4
33	Pre	0	3	4	7	6	5	5	5	4	4
34	Pre	1	2	5	5	6	5	5	5	5	5
35	Pre	0	1	1	3	3	5	5	5	5	4
36	Pre	0	1	2	2	4	5	5	5	5	4
37	Pre	0	1	3	4	6	5	5	5	5	4
38	Pre	0	1	1	3	4	5	5	5	4	4
39	Pre	0	1	3	4	3	5	5	5	5	5
40	Pre	0	1	5	6	7	5	5	5	5	5
41	Pre	0	5	6	8	7	5	5	5	5	5
42	Pre	0	1	2	4	4	5	5	5	5	5
43	Pre	0	1	4	4	6	5	5	4	4	4
44	Pre	0	2	3	5	4	5	5	5	5	5
45	Pre	0	1	4	4	6	5	5	5	5	4
46	Pre	0	5	7	7	8	5	5	5	5	5
47	Pre	0	1	3	3	3	5	5	5	4	4
48	Pre	0	3	6	6	8	5	5	5	5	5
49	Pre	0	2	6	7	6	5	5	5	5	5
50	Pre	0	6	8	8	9	5	5	5	4	4

		Accent Ratings (Scale: 0-9)					Intelligibility Ratings (Scale: 1-5)				
Lstnr ID	Pre/Post	N	M	MM	MS	S	N	M	MM	MS	S
Min	Pre	0	1	1	2	3	5	4	4	3	2
Max	Pre	1	6	8	8	9	5	5	5	5	5
Mean	Pre	0.02	2.24	4.08	4.90	5.44	5	4.98	4.94	4.72	4.34
Var		0.02	1.78	2.77	2.95	2.41	0.0	0.02	0.06	0.25	0.43
Std Dev		0.14	1.33	1.66	1.72	1.55	0.0	0.14	0.24	0.50	0.66

		Accent Ratings (Scale 0-9)					Intelligibility Ratings (Scale: 1-5)				
Lstnr ID	Pre/Post	N	M	MM	MS	S	N	M	MM	MS	S
1	Post	0	3	3	3	4	5	5	5	5	4
2	Post	0	3	4	4	5	5	5	5	5	5
3	Post	0	2	1	6	6	5	4	5	4	4
4	Post	0	3	5	6	7	5	5	4	4	3
5	Post	0	3	6	4	7	5	5	5	5	4
6	Post	0	4	5	5	7	5	5	5	4	4
7	Post	0	7	5	4	7	5	5	5	5	5
8	Post	0	7	6	8	7	5	4	5	4	4
9	Post	0	3	4	5	5	5	5	5	5	4
10	Post	0	2	4	4	6	5	5	5	5	5
11	Post	0	3	2	3	5	5	5	5	5	5
12	Post	0	2	5	6	8	5	5	5	5	4
13	Post	0	2	2	3	4	5	5	4	4	4
14	Post	0	4	3	6	7	5	5	5	5	5
15	Post	0	2	3	5	5	5	5	5	4	4
16	Post	0	2	6	6	6	5	5	5	5	5
17	Post	0	4	6	8	8	5	5	5	5	5
18	Post	0	5	6	6	8	5	5	5	5	4
19	Post	0	5	5	7	6	5	5	4	4	3
20	Post	0	5	6	6	6	5	5	5	5	4
21	Post	0	2	3	4	6	5	5	5	5	5
22	Post	0	4	6	6	7	5	5	5	5	4
23	Post	0	4	6	4	7	5	5	5	5	5
24	Post	0	5	5	6	7	5	5	5	4	5
25	Post	0	2	5	3	6	5	5	5	5	4
26	Post	0	2	3	4	7	5	5	5	4	4
27	Post	0	1	5	6	7	5	5	5	5	5
28	Post	0	5	6	6	8	5	5	5	5	5
29	Post	0	4	6	7	6	5	5	5	4	4
30	Post	0	4	3	7	7	5	5	5	5	5
31	Post	0	4	3	4	7	5	5	5	5	5
32	Post	0	2	2	3	7	5	5	5	5	4
33	Post	0	4	6	8	7	5	5	5	4	4
34	Post	0	2	5	5	6	5	5	5	5	5
35	Post	0	1	1	3	3	5	5	5	5	4
36	Post	0	1	3	3	4	5	5	5	5	5
37	Post	0	2	4	5	7	5	5	5	5	4
38	Post	0	3	2	3	4	5	5	5	4	4
39	Post	0	2	1	4	4	5	5	5	5	5
40	Post	0	1	6	7	7	5	5	5	5	5
41	Post	0	5	3	6	7	5	5	5	5	4
42	Post	0	1	2	4	4	5	5	5	5	5
43	Post	0	2	4	6	6	5	5	4	4	4
44	Post	0	3	2	5	6	5	5	5	5	4
45	Post	0	2	5	6	7	5	5	5	5	5
46	Post	0	4	7	6	8	5	5	5	5	5
47	Post	0	1	2	4	4	5	5	5	4	4
48	Post	0	7	6	8	9	5	5	5	5	5
49	Post	0	2	3	7	7	5	5	5	5	5
50	Post	0	5	8	8	9	5	5	5	5	4
Min	Post	0	1	1	3	3	5	4	4	4	3
Max	Post	0	7	8	8	9	5	5	5	5	5
Mean	Post	0	3.16	4.20	5.26	6.30	5	4.96	4.92	4.72	4.42
Var		0	2.55	3.06	2.44	1.93	0	0.04	0.08	0.21	0.33
Std Dev		0	1.60	1.75	1.56	1.39	0	0.20	0.27	0.45	0.57

REFERENCES

Akahane-Yamada, R., & Tohkura, Y.I. (1997, November). *Effects of audio and audio/visual training on speech communication by non-native listeners*. Paper presented at the International Symposium on Speech Perception by Non-Native Listeners, Boston, MA.

Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmentals. *System, 20,* 51-62.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between NS judgments of NN pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42,* 529-555.

Arends, N., Povel, D.J., Van Os, E., & Speth, L. (1990). Predicting voice quality of deaf speakers on the basis of glottal characteristics. *Journal of Speech and Hearing Research, 33*, 116-122.

Arslan L., & Hansen J. (1997). A study of temporal features and frequency characteristics in American English foreign accent. *Journal of Acoustical Society of America, 102*, 28-40.

Aviation Safety Reporting System (1997). An analysis of international airspace incidents. (Quick Response No. 290). Moffett Field, CA: National Aeronautics and Space Administration.

Barefoot, S., Bochner, J., Johnson, B., & von Eigen, B. (1993). Rating deaf speakers' comprehensibility: An exploratory investigation. *American Journal of Speech-Language Pathology, 2,* 31-35.

Beddor, P.S., & Gottfried T.L. (1995). Methodological issues in cross-language speech perception research with adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 207-232). Baltimore, MD: York Press.

Bergman, M. (1980). *Aging and the perception of speech.* Baltimore, MD: University Park Press.

Beukelman, D.R., & Yorkston, K.M. (1979). The relationship between information transfer and speech intelligibility of dysarthric speakers. *Journal of Communication Disorders, 12,* 189-196.

Beukelman, D.R., & Yorkston, K.M. (1980). The influence of passage familiarity on intelligibility estimates of dysarthric speech. *Journal of Communication Disorders, 13,* 33-41.

Bilger, R.C., Nuetzel, J. M., Rabinowitz W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research, 27,* 32-48.

Brodkey, D. (1972). Dictation as a measure of mutual intelligibility: A pilot study. *Language Learning, 22,* 203-220.

Buus, S., Florentine, M., Scharf, B., & Canevet G. (1986). Native, French listeners' perception of American-English in Noise. *Proceedings of Inter-Noise''86,* 895-898.

Catford, J.D. (1950). Intelligibility. *English Language Teaching, 5,* 7-15.

Celce-Murcia, M., Brinton, D.M., & Goodwin, J.M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages.* Cambridge: University Press.

Chun, D.M. (1989). Teaching tone and intonation with microcomputers. *CALICO Journal,* 21-46.

Comrie, B. (1987). *The world's major languages.* New York: Oxford University Press.

Denes, P.B., & Pinson, E.N. (1993). *The speech chain: The physics and biology of spoken language* (2nd ed.). New York: W.H. Freeman.

Dirks, D.D., Kamm, C.A., Dubno, J.R., & Velde, T.M. (1981). Speech recognition performance at loudness discomfort levels. *Scandinavian Audiology, 10,* 239-246.

Eisling, J., & Wong, R. (1983). Voice quality settings and the teaching of pronunciation. *TESOL Quarterly, 17,* 89-94.

Elliott, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *Journal of the Acoustical Society of America, 22,* 6-13.

Fayer, J.M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning, 37*, 313-326.

Flege, J.E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America, 84*, 70-79.

Flege, J.E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-272). Baltimore, MD: York Press.

Flege, J.E., & Fletcher, K.L., (1992). Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustic Society of America, 9*1, 370-389.

Flege, J., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America, 76,* 708-721.

Flege, J.E., Munro, M. J, & MacKay, I.R.A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97*, 3125-3134.

Fletcher, H. (1953). *Speech and hearing in communication.* Princeton, NJ: Van Nostrand.

Florentine, M. (1985). Non-native listeners' perception of American-English in noise. *Proceedings of Inter-noise '85,* 1021-2024.

Florentine, M., Buus, S., Scharf, B., & Canevet, G. (1984). Speech perception thresholds in noise for native and non-native listeners. *Journal of the Acoustical Society of America, 75*, Suppl. 1, s84.

Gagne, J.P. (1994). Visual and audiovisual speech perception training: Basic and applied research needs. [Monograph]. *Journal of the Academy of Rehabilitative Audiology,* (Suppl. 27*),* 133-159.

Gallego, J. (1990). The intelligibility of three nonnative English-speaking teaching assistants: An analysis of student-reported communication breakdowns. *Issues in Applied Linguistics, 1,* 219-237.

Gass, S. & Varonis, E. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning, 34,* 65-89.

Griffiths, J.D. (1967). Rhyming minimal contrasts: A simplified diagnostic articulation test. *Journal of the Acoustical Society of America, 42,* 236-241.

Hinofotis, F., & Bailey K. (1980). American undergraduate reactions to the communication skills for foreign teaching assistants. In J. Fisher, M. Clarke, & J. Schacter (Eds.), *On TESOL '80: Building bridges* (pp. 120-133). Washington, DC: TESOL.

Hutcherson, R. W., Dirks, D.D., & Morgan, D.E. (1979). Evaluation of the speech perception in noise (SPIN) test. *Otolaryngology Head Neck Surgery, 87,* 239-245.

Imaizumi, S. (1997, November). *Neural processes of audio-visual L2 speech sounds.* Paper presented at the International Symposium on Speech Perception by Non-Native Listeners, Boston, MA.

Johansson, S. (1978). *Studies of error gravity: Native reactions to errors produced by Swedish learners of English.* Göteborg, Sweden: Acta Universitatis Gothoburgensis.

Kalikow, D., Stevens, K., & Elliott, L. (1977). Development of a test of speech intelligibility in noise using sentence material with controlled word predictability. *Journal of the Acoustical Society of America, 61,* 1337-1351.

Kent, R.D. (Ed.) (1992). *Intelligibility in speech disorders: Theory, measurement, and management.* Amsterdam: John Benjamins.

Kent, R.D. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology, 5,* 7-23.

Kent, R.D. (Ed.) (1997). *The speech sciences.* San Diego: Singular Publishing Group.

Kent, R.D., Kim, H., Weismer, G., Kent, J.F., Rosenbek, J.C., Brooks, B.R., & Workinger, M. (1994). Laryngeal dysfunction in neurological disease: Amyotrophic lateral sclerosis, Parkinson's disease, and stroke. *Journal of Medical Speech-Language Pathology, 2,* 157-175.

Kent, R.D., & Read, W. C. (1992). *The acoustic analysis of speech.* San Diego: Singular Publishing Group.

Kent, R.D., Weismer, G., Kent, J.F., & Rosenbeck, J.C. (1989). Toward explanatory intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54,* 482-499.

Kreiman, J., Gerratt, B.R., Kempster, G.B., Erman, A., & Berke, G. (1993). Perceptual evaluation of voice quality: Review, tutorial and a framework for future research. *Journal of Speech and Hearing Research, 36*, 21-40.

Kreiman, J., Gerratt, B.R., Precoda, K., & Berke, G. (1992). Individual differences in voice quality perception. *Journal of Speech and Hearing Research, 35*, 512-520.

Kuhl, P.K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect." In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121-154). Baltimore, MD: York Press.

Lane, H. (1963). Foreign accent and speech distortion. *Journal of the Acoustical Society of America, 35,* 451-453.

Language barriers can cost lives. (1996, November 16). *Citrus County Chronicle*, p. A10.

Major, R.C. (1987). Measuring pronunciation accuracy using computerized techniques. *Language Testing, 4*, 155-169.

Maxwell, D.L., & Satake E. (1997). *Research and statistical methods in communication disorders.* Baltimore, MD: Williams & Wilkins.

Mayo, L.H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research, 40*, 686-693.

Metz, D.E., Samar, V.J., Schiavetti, N., & Sitler, R.W. (1990). Acoustic dimensions of hearing-impaired speakers' intelligibility: Segmental and suprasegmental characteristics. *Journal of Speech and Hearing Research, 33*, 476-487.

Metz, D.E., Samar, V.J., Schiavetti, N. Sitler, R.W., & Whitehead, R.L. (1985). Acoustic dimensions of hearing-impaired speakers' intelligibility. *Journal of Speech & Hearing Research, 28,* 345-355.

Metz, D.E., Schiavetti, N., & Sitler, R.W. (1980). Toward an objective description of the dependent and independent variables associated with intelligibility assessments of hearing impaired adults. In J. Subtelny (ed.), *Speech assessment and speech improvement for the hearing impaired*, 72-81. Washington, DC: A.G. Bell Association for the Deaf.

Monsen, R.B. (1978). Toward measuring how well hearing-impaired children speak. *Journal of Speech and Hearing Research, 21,* 197-219.

Morgan, D.E., Kamm, C.A., & Velde, T.M. (1981). Form equivalence of the speech perception in noise (SPIN) test. *Journal of the Acoustical Society of America, 69,* 1791-1798.

Morley, J. (1993). The pronunciation component in teaching English to speakers of other languages. In S. Silberstein (Ed.). *State of the art TESOL essays: Celebrating 25 years of the discipline* (pp. 310-349). Washington, DC: TESOL.

Munro, M.J. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition, 17,* 17-34.

Munro, M., & Derwing, T. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45,* 73-97.

Munro, M., & Derwing, T. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38,* 289-306.

Nabelek, A.K., & Donahue, A.M. (1984). Perception of consonants in reverberation by native and non-native listeners. *Journal of the Acoustical Society of America, 75,* 632-634.

Nash, R. (1972). Phonemic and prosodic interference and intelligibility. In A. Rigault & R. Charbonneau (Eds.), *Proceedings of the Seventh International Congress of Phonetic Sciences* (pp. 570-573). The Hague: Mouton.

Nickerson, R.S., & Stevens, K.N. (1980). Approaches to the study of the relationship between intelligibility and physical properties of speech. In J. Subtelny (Ed.), *Speech assessment and improvement for the hearing impaired* (pp. 338-364). Washington, DC: AG Bell Association for the Deaf.

Nicolosi, L., Harryman, E., & Kresheck, J. (1989). *Terminology of communication disorders: Speech-language-hearing.* (3rd ed.). Baltimore, MD: Williams & Wilkins.

Orlikoff, R.F., & Baken, R.J. (1993). *Clinical speech and voice measurement.* San Diego: Singular Publishing Group.

Owen, J.H. (1981). Influence of acoustical and linguistic factors on the SPIN test difference score. *Journal of the Acoustical Society of America, 70*, 678-682.

Palmer, J. (1976). Linguistic accuracy and intelligibility. In *Proceedings of the 4th International Congress of Applied Linguistics* (pp. 505-513). Stuttgart, Germany: Hochschul Verlag.

Patkowski, M.S. (1994). The critical age hypothesis and interlanguage phonology. In M. Yavas (Ed.), *First and second language phonology* (pp. 205-222). San Diego: Singular Publishing Group.

Penrod, J. P. (1985). Speech discrimination testing. In J. Katz (Ed.), *Handbook of clinical audiology* (3rd ed., pp. 235-255). Baltimore, MD: Williams & Wilkins.

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics, 15*, 37-52.

Purcell, E., & Suter, R. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning, 30,* 271-287.

Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In R.D. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement, and management* (pp. 11-34). Amsterdam: John Benjamins.

Schiavetti, N., Metz, D.E., & Sitler, R.W. (1981). Construct validity of direct magnitude estimation and interval scaling of speech intelligibility: Evidence from a study of the hearing impaired. *Journal of Speech and Hearing Research, 24,* 441-445.

Schiavetti, N., Sitler, R.W., Metz, D.E., & Houde, R.A. (1984). Prediction of contextual speech intelligibility from isolated word intelligibility measures. *Journal of Speech and Hearing Research, 27,* 623-626.

Schmid, P.M., & Yeni-Komshian, G.H. (1999). The effects of speaker accent and target predictability on perception of mispronunciations. *Journal of Speech, Language, and Hearing Research, 42,* 56-64.

Sitler, R.W., Schiavetti, N., & Metz, D.E. (1983). Contextual effects in the measurement of hearing-impaired speakers' intelligibility. *Journal of Speech and Hearing Research, 26,* 30-34.

Smith L., & Nelson C. (1985). International intelligibility of English: Directions and resources. *World Englishes, 4,* 333-342.

Stansfield, C. W., & Ballard, R.J. (1984). Two instruments for assessing the oral English proficiency of foreign teaching assistants. *Foreign Teaching Assistants,* 101-109.

Stevens, S.S. (1975). *Psychophysics*. New York: Wiley.

Strange, W. (Ed.). (1995). Cross-language studies of speech perception: A historical review. In, *Speech perception and linguistic experience: Issues in cross-language research.* (pp. 3-45). Baltimore, MD: York Press.

Strange, W. (1997, November). *Perception of Non-Native Speech Sounds by Adults.* Paper presented at International Symposium on Speech Perception by Non-Native Listeners conducted at the American Speech, Language and Hearing Convention, Boston, MA.

Strevens, P. (1988). ESP after twenty years: A re-appraisal. In M.L. Tickoo (Ed.), *ESP: State of the art* (pp. 1-13). Singapore: SEAMEO Regional Language Centre.

Subtelny, J. (1977). Assessment of speech with implications for training. In F. Bess (Ed.), *Childhood deafness* (p.183). New York: Grune & Stratton.

Subtelny, J., Orlando, N., & Whitehead, R. (1981). *Speech and voice characteristics of the deaf.* Washington, DC: Alexander Graham Bell Association for the Deaf.

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning, 4,* 177-204.

Tyler, R.S. (1994). The use of speech-perception tests in audiological rehabilitation: current and future research needs. [Monograph]. *Journal of the Academy of Rehabilitative Audiology (Suppl. 27),* 47-66.

Varonis, E., & Gass, S. (1982). The comprehensibility of nonnative speech. *Studies in Second Language Acquisition, 4,* 114-136.

Weismer, G., & Martin, R.E. (1992). Acoustic and perceptual approaches to the study of intelligibility. In R.D. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement, and management* (pp. 67-118). Amsterdam: John Benjamins.

Weismer, G., Kent, R.D., Hodge, M., & Martin, R.E. (1988). The acoustic signature for intelligibility test words. *Journal of the Acoustical Society of America, 84,* 1281-1291.

Werker, J.F. (1986). The effect of multilingualism on phonetic perceptual flexibility. *Applied Psycholinguistics, 7,* 141-156.

Wingstedt, M., & Schulman, R. (1987). Comprehension of foreign accents. In W. Dressler (Ed.), *Phonologica 1984: Proceedings of The Fifth International Phonology Meeting, Eisenstadt* (pp. 339-344). Cambridge: Cambridge University Press.

Yorkston, K.M., & Beukelman, D.R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders, 11,* 499-512.

BIOGRAPHICAL SKETCH

Major Kimberly R. Scott was born on September 28, 1961, in Tucson, Arizona. When she was very young her family moved to Minnesota, where she lived until she graduated from high school in 1979. She received a Bachelor of Science degree in communication disorders from the University of Wyoming in 1983. She graduated from the University of Arkansas for Medical Sciences in 1985 with a Master of Science degree in speech-language pathology. Following graduation, she was employed as a speech-language pathologist in private practice and in a rehabilitation hospital setting in Idaho. In 1987 she joined the United States Air Force as a clinical speech-language pathologist. She served in various medical settings in the United States, the United Kingdom, Germany, and Turkey. Kim was selected by the Air Force Institute of Technology to pursue a Doctor of Philosophy in communication sciences and disorders while remaining on active duty. In August of 1996, she enrolled in the doctoral program at the University of Florida. Her Ph.D. in communication sciences was awarded in August 1999.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Alice M. Dyson

Alice M. Dyson, Chair
Associate Professor of Communication Sciences and Disorders

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Patricia A. Kricos

Patricia A. Kricos
Professor of Communication Sciences and Disorders

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Howard B. Rothman

Howard B. Rothman
Professor of Communication Sciences and Disorders

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Christine M. Sapienza
Associate Professor of Communication Sciences and Disorders

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Jean Casagrande
Professor of Romance Languages and Literatures